MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 0022
De.2. **Measure Title:** Use of High-Risk Medications in the Elderly (DAE)
Co.1.1. **Measure Steward:** National Committee for Quality Assurance
De.3. **Brief Description of Measure:** There are two rates for this measure:

- The percentage of patients 65 years of age and older who received at least one high-risk medication.
- The percentage of patients 65 years of age and older who received at least two prescriptions for the same high-risk medication.

For both rates, a lower rate represents better performance.
1b.1. **Developer Rationale:** Lowering the use of high-risk medications in the elderly population should decrease morbidity and mortality associated with adverse drug reactions.

S.4. **Numerator Statement:** Numerator 1: Patients who received at least one high-risk medication during the measurement year.

Numerator 2: Patients who received at least two prescriptions for the same high-risk medication during the measurement year.

For both numerators a lower rate indicates better performance.
S.7. **Denominator Statement:** All patients 65 years of age and older.
S.10. **Denominator Exclusions:** Patients who were enrolled in hospice care at any time during the measurement year.

De.1. **Measure Type:** Process
S.23. **Data Source:** Administrative claims, Electronic Clinical Data, Electronic Clinical Data : Pharmacy
S.26. **Level of Analysis:** Health Plan, Integrated Delivery System

IF Endorsement Maintenance – Original Endorsement Date: Aug 10, 2009 **Most Recent Endorsement Date:** Aug 09, 2012

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

Maintenance of Endorsement -- Preliminary Analysis

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

Criteria 1: Importance to Measure and Report

1a. Evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a *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.

The developer provides the following evidence for this measure:

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

Evidence Summary

- This measure assesses if older adults are prescribed medications that are potentially inappropriate. The measure is directly based on specific recommendations in the American Geriatrics Society (AGS) Beers Criteria identifying which drugs are potentially inappropriate for all older adults.

Changes to evidence from last review

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

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

Updates:

- The American Geriatrics Society 2015 Beers Criteria Update Expert Panel graded the evidence. The evidence had not been graded the last time the measure was submitted for maintenance. The 2015 review by the AGS 2015 Beers Criteria Update Expert Panel, which this measure is based on, included review of 60 systematic reviews and meta analyses, 49 randomized control trials (RTCs) and 233 observational studies and other types of publications. Overall, the quality of the evidence for each of the medications included in the Beers Criteria recommendations is good.

Exception to evidence

N/A

Guidance from the Evidence Algorithm

1-No→3-Yes → 4-Yes→ 5a-HIGH

Questions for the Committee:

If the developer provided updated evidence for this measure:

- *Questions specific to the measure information provided on evidence*
- *For process 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?*

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

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

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

- The average performance for the first rate (at least one high-risk medication) has decreased from 21.0% in 2012 to 13.2%.
- The average performance for the second rate (dispensing two different high-risk medications) has decreased

from 6.5% in 2012 to 2.1% in 2014. In 2014, for both populations the eligible population was 22,043.

- Overall, the gap seems to be closing overtime and there is still opportunity for improvement. The developers have updated the specifications to include multiple prescribing events for the same high-risk medication. They expect the new rate will be higher.

Disparities

- The developer summarized results from a retrospective cohort study of 966,000 men and women treated by the Veteran's Health Administration. The study showed that women were more likely than men to receive medications that may have harmful interactions with chronic conditions as described by the Beers Criteria.

Questions for the Committee:

- *Specific question on information provided for gap in care.*
- *Is there a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*

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

Comments: ** measure is for patients >65 who are prescribed at least one medication from a list of meds which are recognized to potentially cause serious drug events in the elderly.

** This is a process measure with two rates. The measure assesses if older adults are prescribed medications that are potentially inappropriate. It is based on specific recommendations in the Beers Criteria. Updated evidence for 2015 was provided.

1b. Performance Gap

Comments: ** Yes. There is a significant opportunity nationally to reduce adverse medication events in the elderly. For this specific metric, which has been used as a HEDIS measure, there is both a gap from 10th percentile to 90th percentile, as well as a demonstrated gap in overall performance, although this has improved since 2012.

** While the average performance has improved between 2012 and 2014, there is still a large difference in performance between the plans at the 90th and 10th percentiles which represents a gap in care. Disparities data is not available in NCQA, but there is some evidence that women are more likely to receive a potentially inappropriate medication than men, based on a retrospective cohort study in the VHA.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Administrative claims

Specifications:

- This measure includes two rates:
  - Rate 1: The percentage of patients 65 years of age and older who received at least **one** high-risk medication.
  - Rate 2: The percentage of patients 65 years of age and older who received at least **two** prescriptions for

the same high-risk medication.

- The measure uses administrative claims data (including pharmacy claims) to assess the percentage of patients age 65 and older who have been prescribed a high-risk medication.
  - Rate 1 identifies:
    - Patients who have at least one dispensing event (for any length of time) for one of the medications in table DAE-A;
    - Patients who have at least one dispensing event for a medication in Table DAE-B where days supply exceeds the days supply criteria listed for the medication;
    - Patients who have at least one dispensing event for a medication in Table DAE-C where average daily dose exceeds the average daily dose criteria listed for the medication.
  - Rate 2 identifies:
    - Patients who have at least two dispensing events (for any length of time) on different dates of service for the same medication (as defined in table DAE-A);
    - Patients who have at least two dispensing events (for any length of time) on different dates of service for medications in the same class (as defined in Table DAE-B);
    - Patients who have at least two dispensing events for a medication in Table DAE-C where average daily dose exceeds the average daily dose criteria listed for the medication.

Questions for the Committee :

- *Specific questions on the specifications, codes, definitions, etc.*
- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

Describe any updates to testing

- Prior testing was conducted on HEDIS Health Plan performance data from 2010; for this maintenance review, testing was updated using HEDIS Health Plan performance data from 2012-2014.

SUMMARY OF TESTING

Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing

- Reliability testing was conducted on HEDIS Health Plan performance data from 2012-2014; the developers do not provide details on the size or characteristics of the test population.
- The developer conducted a signal-to-noise analysis of the measure score using a beta binomial method.
- As described by the developer, a signal-to-noise analysis assesses reliability of a measure by estimating the proportion of variability in measured performance that can be explained by real differences in performance. A reliability of zero implies that all the variability in a measure is attributable to measurement error; a reliability of one implies that all the variability is attributable to real differences in performance.

Results of reliability testing

- The developer reports that reliability for this measure was calculated as **0.99814** for receipt of one or more high-risk prescriptions and **0.99594** for receipt of two or more high-risk prescriptions.
- The developer states that a reliability score greater than or equal to 0.7 is considered very good, and suggests that testing results demonstrate that both indicators in this measure are highly reliable.

Questions for the Committee:

○ *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

Guidance from the Reliability Algorithm

[Box 1] Specifications precise and unambiguous → [Box 2] Empirical testing conducted on the measure as specified → [Box 4] Testing conducted at the measure score level → [Box 5] → Testing method described and appropriate → [Box 6] High certainty or confidence that measure scores are reliable → [Box 6a]

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

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**
Specification not completely consistent with evidence

Question for the Committee:

○ *Are the specifications consistent with the evidence?*

2b2. Validity testing

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, summarize the validity testing from the prior review:

- For prior endorsement reviews, the developer reported only that face validity had been affirmed by an Expert Panel.

Describe any updates to validity testing:

- For this maintenance review, the developer has provided the results of empirical validity testing and additional details of the face validity assessment.

SUMMARY OF TESTING
Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:
☐ **Face validity only**
☒ **Empirical validity testing of the measure score**

Validity testing method:
- The developer tested the measure for construct validity by exploring whether the two rates within this measure were correlated with each other and with another measure of medication safety (Potentially Harmful Drug-Disease Interactions in the Elderly).
- The developer hypothesized that organizations that perform well on one of the indicators should perform well on the other indicator as well as the other medication safety measure.

- The correlations were assessed using a Pearson correlation test.
- The developer explains that a Pearson correlation test estimates the strength of the linear association between two continuous variables; the magnitude of correlation ranges from -1 and +1.
    - A value of 1 indicates a perfect linear dependence in which increasing values on one variable is associated with increasing values of the second variable. A value of 0 indicates no linear association. A value of -1 indicates a perfect linear relationship in which increasing values of the first variable is associated with decreasing values of the second variable.
- In addition, the developer used two expert panels to assess the measure's face validity.

Validity testing results:

- The developer provides the following table reporting the results of the Pearson correlation test:

Table 1a. Correlations among both rates in the measure and a drug-disease interaction measure[1]

Measure	Pearson Correlation Coefficients	
	Rate 1: One high-risk medication	Rate 2: Two high-risk medications
Rate 1: One high-risk medication		
Rate 2: Two high-risk medications	.8745	
Drug-disease interaction: History of Falls	0.307	.2735
Drug-disease interaction: Dementia	0.454	.4390
Drug-disease interaction: Chronic Kidney Disease	0.367	.3552
Drug-disease interaction: Total	0.386	.3913

Note: All correlations are significant at p<.05

[1]The *Potentially Harmful Drug-Disease Interactions in the Elderly* measure has four rates. The first rate assesses the percentage of patients 65 and older with a history of falls who received a high-risk medication. The second rate assesses the percentage of patients 65 and older with dementia who received a high-risk medication. The third rate assesses the percentage of patients 65 and older with chronic kidney disease who received a high-risk medication. The fourth rate is the sum of the three numerators divided by the sum of the three denominators for the three previous rates. Note: "high-risk" medications for each condition are based on recommendations in Table 3 of the American Geriatrics Society Beers Criteria.

- The developer states that Pearson correlation coefficients with absolute value of less than 0.3 are generally considered indicative of weak associations whereas absolute values of 0.3 or higher denote moderate to strong associations.
- The developer's interpretation of these results is that they confirm the hypothesis that rates in the measure are correlated with each other as well as with another measure of medication safety, suggesting they represent the same underlying quality construct of prescribing inappropriate medications for patients with the corresponding illnesses.
- Regarding the face validity assessment, the developer reports that the measure was deemed to have the desirable attributes of a HEDIS measure in 2006 (relevance, scientific soundness, and feasibility).
- The developer suggests that these results indicate the expert panels showed agreement that the measures as specified will accurately differentiate quality across health plans.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The developer reports that this measure has only one exclusion for individuals who are in hospice during the measurement year.

- The developer does not provide analysis or a rationale to support exclusions.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

Conceptual rationale for SDS factors included ? ☐ **Yes** ☒ **No**

SDS factors included in risk model? ☐ **Yes** ☒ **No**

Risk adjustment summary

- The developer states that, as a measure assessing the use of high-risk medication in a general elderly population, risk adjustment is not indicated.

Questions for the Committee:

- *Is it appropriate to not risk-adjust this measure?*
- *Do you agree with the developer's rationale that there is no conceptual basis for adjusting this measure for SDS factors?*
- *Do you agree with the developer's decision, based on their analysis, to not include SDS factors in their risk-adjustment model?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- To demonstrate that meaningful differences in performance can be identified, the developer provides comparison of means and percentiles from 2012 to 2014 HEDIS Health Plan Performance Data.
- The developer notes that if sample size is >400, they use an analysis of variance against established benchmarks.
- The following two tables are provided to present the results of this analysis:

At least one high-risk prescription

	Number of Plans	Mean	Standard Deviation	Min	Max	10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile
2012	498	21.0	6.4	5.5	54.6	14.0	16.5	19.9	24.5	30.0
2013	494	18.0	6.1	1.0	50.5	11.5	13.8	16.7	21.1	25.8
2014	488	13.2	6.0	2.6	46.8	7.6	9.2	11.6	16.1	21.7

At least two high-risk prescriptions

	Number of Plans	Mean	Standard Deviation	Min	Max	10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile
2012	498	6.5	2.9	1.2	25.2	3.5	4.7	6.0	7.8	10.1
2013	494	3.1	2.3	0.0	20.6	1.1	1.7	2.4	4.0	6.0
2014	488	2.1	2.0	0.0	20.8	0.6	0.9	1.4	2.5	4.6

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- The developer reports that this question is inapplicable, as only one data source is used to calculate the measure (pharmacy claims).

2b7. Missing Data

N/A

Guidance from the Validity Algorithm

[Box 1] Specifications consistent with evidence → [Box 2] Potential threats to validity addressed (though little support provided for exclusions, missing data) → [Box 3] Empirical validity testing conducted using the measure as specified → [Box 6] Testing conducted at the measure score level → [Box 7] Testing method described and appropriate → [Box 8] Moderate certainty or confidence that measure scores are reliable → [Box 8b]

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

Committee pre-evaluation comments

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

2a1. & 2b1. Specifications

Comments: ** This measure affects large numbers of people and is a leading cause of morbidity and mortality. Also involves high resource use and has consequences for society.

2a2. Reliability Testing
Comments: ** Reliability testing .99882 for 1 Rx and .99819 for 2 Rx demonstrating high reliability.
Validity - performed face validity with 2 panels of experts plus public comment as well as construct validity using Pearson correlation of the 2014 HEDIS scores from 488 medicare health plans with an average population of over 22,000
** Specifications are clear and consistent with evidence.
** Yes for demonstrated reliability
** Reliability used is signal-to-noise ratio, calculated at .099814 and 0.99594, suggesting high reliability.

2b2. Validity Testing
Comments: ** Yes for validity
** The measure was tested for face validity using two expert panels (GMAP and NCQA's performance measurement committee) There was a high correlation between the first and second rate in the measure. There were moderate correlations between both rates and the four rates in the other medication safety measure.

2b3. Exclusions Analysis
2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures
2b5. Identification of Statistically Significant & Meaningful Differences In Performance
2b6. Comparability of Performance Scores When More Than One Set of Specifications
2b7. Missing Data Analysis and Minimizing Bias
Comments: ** exclusions were only patients enrolled in hospice for the reporting year.
**This is not an eMeasure.
**The only exclusion is for individuals on hospice, which is supported.
**Risk adjustment is not indicated
**Analysis shows the methods for scoring and analysis allow for identification of statistically significant and clinically meaningful differences.

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.

- This measure is generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score)
- 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.
- ALL data elements are in defined fields in a combination of electronic sources
- The developer developed a precise, standardized methodology for verifying the integrity of HEDIS collection and calculation processes through a two-part program consisting of an overall information systems capabilities assessment followed by an evaluation of the MCO´s ability to comply with HEDIS specifications. In which certified auditors using standard audit methodologies will help enable purchasers to make more reliable "apples-to-apples" comparisons between health plans.
- In addition to the HEDIS Audit, the developer provides a system to allow "real-time" feedback from measure users. The Policy Clarification Support System receives thousands of inquiries each year on over 100 measures.

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

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: ** Data is routinely generated during care delivery and is available in EHR or other electronic sources
** All data elements are in electronic sources

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

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

Current uses of the measure
Publicly reported? ☒ **Yes** ☐ **No**

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

Accountability program details:

- HEALTH PLAN RATINGS/REPORT CARDS: This measure is used to calculate health plan ratings which are reported in Consumer Reports and on the NCQA website. These ratings are based on performance on HEDIS measures

among other factors. In 2012, a total of 455 Medicare Advantage health plans, 404 commercial health plans and 136 Medicaid health plans across 50 states were included in the ratings. In 2015 NCQA announced a change in methodology and changed Health Plan Rankings to Health Plan Ratings.

- CMS EHR INCENTIVE PROGRAM (MEANINGFUL USE): The Medicare and Medicaid Electronic Health Care Record (EHR) Incentive Programs provide incentive payments to eligible professionals, eligible hospitals, and critical access hospitals (CAHs) as they adopt, implement, upgrade or demonstrate meaningful use of certified EHR technology.
- CMS Medicare Part D: This measure is aligned with the Pharmacy Quality Alliance's Use of High-Risk Medications in the Elderly measure which is reported by Medicare Part D plans. Organizations contracted to offer Medicare Part D benefits are required to report data to CMS on a variety of measures. CMS has developed reporting standards and data validation specifications with respect to the Part D reporting requirements. These standards and specifications provide a review process for Medicare Advantage Organizations (MAOs), Cost Plans, and Part D sponsors to use to conduct data validation checks on their reported Part D data. The data validation is "retrospective," referring to the fact that it normally occurs in the year subsequent to the measurement year.
- STATE OF HEALTH CARE ANNUAL REPORT: This measure is publically reported nationally and by geographic regions in the NCQA State of Health Care annual report. This annual report published by NCQA summarizes findings on quality of care. In 2012 the report included measures on 11.5 million Medicare Advantage beneficiaries in 455 Medicare Advantage health plans, 99.4 million members in 404 commercial health plans, and 14.3 million Medicaid beneficiaries in 136 plans across 50 states.
- HEALTH PLAN ACCREDITATION: This measure is used in scoring for accreditation of Medicare Advantage Health Plans. In 2012, a total of 170 Medicare Advantage health plans were accredited using this measure among others covering 7.1 million Medicare beneficiaries. Health plans are scored based on performance compared to benchmarks.
- HEDIS ACCOUNTABLE CARE ORGANIZATION ACCREDITATION: This measure is used in NCQA's ACO Accreditation program, that helps health care organizations demonstrate their ability to improve quality, reduce costs and coordinate patient care. ACO standards and guidelines incorporate whole-person care coordination throughout the health care system.
- HEDIS PHYSICIAN ACCREDITATION: This measure is used in NCQA's Physician Accreditation program, that helps physicians demonstrate their ability to improve quality, reduce costs and coordinate patient care.
- PHYSICIAN QUALITY REPORTING SYSTEM: This measure is used in the Physician Quality Reporting System (PQRS) which is a reporting program that uses a combination of incentive payments and payment adjustments to promote reporting of quality information by eligible professionals (EPs). Eligible professionals who satisfactorily report data on quality measures for covered Physician Fee Schedule services furnished to Medicare Part B beneficiaries (including Railroad Retirement Board and Medicare Secondary Payer) receive these payment incentives and adjustments.

Improvement results:

- Over the past three years the two rates in this measure have shown approximately 8% improvement across health plans in the first rate and 4.4% in the second rate.
- In an effort to see greater room for improvement, the developer changed the rate to assess multiple dispensing events for the same high-risk medication.
- Due to recent updates to the medications included in the measure, future rates may show greater room for improvement and variation in performance.

Unexpected findings (positive or negative) during implementation:

- There were no identified unintended consequences for this measure during testing or since implementation.

Potential harms:

- If this measure were to be implemented poorly, there is concern that it could lead to reduced access to medications. There will always be individual cases that will warrant the use of a potentially harmful medication and clinicians should weigh the risks and benefits of using these medications for their individual patients.

Feedback :

- Developer did not identify any specific feedback loops related to this 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

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: ** Currently used for multiple public reporting venues including HEDIS, PQRS, Meaningful Use and NCQA accreditation ** Used in public reporting, health plan ratings, meaningful use, payment programs, accreditation and regulatory programs, QI and benchmarking.

Criterion 5: Related and Competing Measures

Related or competing measures

2993 : Potentially Harmful Drug-Disease Interactions in the Elderly

This measure is being submitted as a new measure for NQF endorsement during this current Patient Safety project.

Harmonization

Measure 2993 and NQF 0022 have a similar focus (measuring potentially inappropriate medication use in the elderly) and reporting level (health plan), however they have different target populations. Measure 2993 targets patients with a specific condition or disease that can experience adverse effects when combined with certain medications that are recommended to be avoided for that condition. This measure (NQF 0022) targets a larger population of all older adults and assesses use of high-risk medications that have been recommended to be avoided in all older adults.

Pre-meeting public and member comments

Submitted by: ADVault, Inc.

ADVault believes that people live better lives and, if in a health crisis, can receive better care when they have confidence they can be involved in the creation and implementation of their medical treatment plans and decisions, factors extremely important when it comes to high risk medication being prescribed to the elderly. To do so, they must be able to communicate and express their goals, preferences and priorities for care in a meaningful and actionable way so providers can consider those thoughts. At some point in life, everyone will lose his or her ability to communicate effectively and understand what is being asked of him or her. Healthcare agents should have the confidence to know those value statements as well, in order to fulfill their role as surrogate decision-makers. Non-surrogate family members are comforted with third-party decision-making if they have proof the patient's voice is being heard, clearly understood, and to the extent possible, honored.

Therefore, ADVault strongly recommends providers (1) search for a person's digital emergency, critical and advance care plan (ECACP) upon admission and each time the patient is transitioned to a new site of care, (2) review and update the ECACP in various stages of a person's admission (outpatient or inpatient) and/or illness to ensure respect for the person's goals, preferences and priorities for care, (3) link the digital ECACP to the EHR

and/or patient portal in order to ease access and address security, privacy and patient consent concerns, (4) track and make available the number of ECACPs found, opened and re-visited, and the impact they have on the care of the patient, as well as patient, family and caregiver satisfaction, such data to be reported in a manner such that: (a) consumers can make better choices about hospitals and doctors; (b) doctors improve the satisfaction and quality of their work; and (c) hospital administrators gauge performance and align caregiving goals with actual outcomes. Finally, if no ECACP can be found via standards-based healthcare IT transport mechanisms, the hospital/provider should engage the patient to create one whenever possible.

Submitted by: Centers for Disease Control and Prevention

CDC strongly supports a patient safety measure related to medication management in older adults; however, we are concerned that the CDC data cited is not appropriately applied and the measure may not efficiently reduce adverse drug events (ADEs). First, the measure rationale is that reduction in "high-risk medication" (HRM) use "should decrease morbidity and mortality" associated with ADEs and CDC data are cited in the discussion of measure impact. However, CDC data indicate the opposite--Beers Criteria (BC) HRMs are not leading causes of emergency department (ED) visits or hospitalizations for ADEs (Ann Intern Med 2007;147:755-65; N Engl J Med 2011;365:2002-12). Approximately 1% of U.S. hospitalizations for ADEs among older adults involve BC HRMs, while approximately 66% involve 3 other drug classes (warfarin, antidiabetics, oral antiplatelets). After accounting for prescribing, the hospitalizations rate for ADEs from these 3 drug classes is at least 40 times higher than the hospitalization rate for ADEs from BC HRMs (N Engl J Med 2011;365:2002-12).
Second, although there are a few studies to support an epidemiologic association of BC HRMs with health outcomes, there are many other studies that do not support this finding. The studies cited in the measure are based on older BC versions. We are not aware of new data demonstrating that use of the updated BC is associated with morbidity, mortality, or resource utilization reductions. Third, using a composite measure targeting hundreds of drugs/interactions obscures the contribution of specific drugs and thus cannot be efficiently used to implement interventions (J Hosp Med 2008;3:87-90). One-half of Medicare Advantage beneficiaries meet criteria for HRM drug-disease interactions, suggesting the measure is not useful for targeting the highest risk drugs. Fourth, basing a broad healthcare quality measure on the "potentially inappropriate" concept is problematic because it supersedes the treating clinician's judgment without having supporting information for that clinical judgment. The 2015 BC update states: "these criteria are not meant to be applied in a punitive manner. Prescribing decisions are not always clear-cut, and clinicians must consider multiple factors...Quality measures must be...measured with limited information and thus...cannot perfectly distinguish appropriate from inappropriate care". The BC is a useful tool to guide individual clinical decisions; however, as a quality measure, it is likely to have minimal population impact. A fundamental criterion of NQF measures is that they be aligned with national health priorities; for medication safety, these have been defined as improving safe use of anticoagulants, antidiabetics, and opioids (health.gov/hcq/ade-action-plan.asp). Incorporation of these medications into national quality measures will go further toward improving health outcomes for older Americans than measures focused on HRMs.

NQF #: 0022	**NQF Project:** Patient Safety Measures-Complications Project

1. IMPACT, OPPORTUITY, EVIDENCE - IMPORTANCE TO MEASURE AND REPORT

Importance to Measure and Report is a threshold criterion that must be met in order to recommend a measure for endorsement. All three subcriteria must be met to pass this criterion. See guidance on evidence.
Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria.
(evaluation criteria)

1c.1 **Structure-Process-Outcome Relationship** *(Briefly state the measure focus, e.g., health outcome, intermediate clinical outcome, process, structure; then identify the appropriate links, e.g., structure-process-health outcome; process- health outcome; intermediate clinical outcome-health outcome)*:
This Patient Safety measure addresses medication management to prevent the harms associated with certain medications in the elderly.

Panels of experts in pharmacology and geriatrics have compiled lists of medications to avoid prescribing for patients 65 years of age or older. The most commonly used list is the Beers criteria, which was introduced in 1991 to serve researchers evaluating prescribing quality in nursing homes. The Beers criteria were updated in 1997 and again in 2003 to include 48 "potentially inappropriate medications" (PIMs) for which, according to the consensus panel, there are more effective or safer alternatives for older patients. (Rothberg)

Reducing prescriptions of high-risk drugs in the elderly also represents an opportunity to reduce the costs associated with the harm from medications (e.g., hospitalizations from drug toxicity) and encourage clinicians to consider safer, alternative medications. Reducing unnecessary prescribing will also help to reduce cost, given that the elderly population represent one third of all prescription drug expenditures in the U.S. but comprises only 13 percent of the population. (Families USA)

While expenditures for prescription drugs in the US are disproportionately clustered among those 65 years and older, (Families USA) this population is twice as likely as those below age 65 to experience adverse drug events and is almost seven times as likely to be hospitalized. (Budnitz 2006) Important factors increasing the risk of adverse drug events in the elderly include prescription of drugs that are generally inappropriate for the elderly, interactions between drugs and pre-existing conditions, and interactions between contra-indicated drugs.

While some drugs are generally appropriate to prescribe in the elderly, the side-effects commonly associated with these drugs pose an extra risk to elderly people with certain pre-existing conditions. For example, the unsteadiness (ataxia) frequently associated with antidepressants may be a particular danger for elderly patients with a history of falls. Clinical guidelines identify drugs that are generally inappropriate for the elderly, as well as drugs that are inappropriate for elderly populations with specific diagnoses or conditions. (Fick)

In 2005, rates of potentially inappropriate medication use in the elderly were as large or larger than in a 1996 national sample, highlighting the need for progress in this area. (Simon)

While some adverse drug events are not preventable, studies estimate that between 30% and 80% of adverse drug events in the elderly are preventable. (MacKinnon)

Reducing the number of inappropriate prescriptions can lead to improved patient safety and significant cost savings. Conservative estimates of extra costs due to potentially inappropriate medications in the elderly average $7.2 billion a year. (Fu)

Medication use by older adults will likely increase further as the U.S. population ages, new drugs are developed, and new therapeutic and preventive uses for medications are discovered. (Rothberg)

By the year 2030, nearly 1 in 5 U.S. residents is expected to be aged 65 years or older; this age group is projected to more than double in number from 38.7 million in 2008 to more than 88.5 million in 2050.1,2 Likewise, the population aged 85 years or older is expected to increase almost 4-fold, from 5.4 million to 19 million between 2008 and 2050.1 As the elderly population continues to grow, the number of older adults who present with multiple medical conditions for which several medications are prescribed continues to increase, resulting in polypharmacy. (Gray)

1c.2-3 Type of Evidence *(Check all that apply)*:
Selected individual studies (rather than entire body of evidence)
Systematic review of body of evidence (other than within guideline development)

Clinical Practice Guideline: *American Geriatrics Society 2015 Updated Beers Criteria for Potentially Inappropriate Medication Use in Older Adults*

1c.4 Directness of Evidence to the Specified Measure *(State the central topic, population, and outcomes addressed in the body of evidence and identify any differences from the measure focus and measure target population)*:
Evidence base is the original Beers study, the Zahn study and the Fick update to the Beers list in 2003.

This measure assesses if older adults are prescribed medications that are potentially inappropriate. The measure is directly based on specific recommendations in the American Geriatrics Society (AGS) Beers Criteria of which drugs are potentially inappropriate for all older adults (i.e., recommendations in Table 2 of the Beers Criteria). The target population in the measure therefore includes all adults age 65 and older. In general, older adults should not receive these medications as the potential harms of their use likely outweigh any benefit.

To translate the Beers Criteria recommendations for use in this quality measure, NCQA uses the following guiding principles to determine which medication classes are included:

1. Include only medications listed in *Table 2: 2015 AGS Beers Criteria for Potentially Inappropriate Medication Use in Older Adults.*
2. Include only prescription medications.
3. Include only medications where the AGS Recommendation indicates "avoid" and the AGS Rationale does not include "avoid for" caveats that cannot be identified from prescription drug claims data.
4. Include medications with caveats only if they can be measured efficiently and reliably from prescription drug claims data.
5. If including a medication in the measure would likely result in the increased use of another potentially harmful medication that is not included in the measure, an exception to these guiding principles may be warranted to reduce this unintended consequence.

1c.5 Quantity of Studies in the Body of Evidence *(Total number of studies, not articles)*: 2

The Beers Criteria was first published in 1991. Since that time the criteria have been regularly updated based off of the existing criteria and any new evidence published since the last update. The AGS forms an expert panel to update the Beers Criteria every few years. The panel works from the previous evidence review and then reviews any new evidence published since that last review to update the recommendations in the Beers Criteria. The 2015 review by the AGS 2015 Beers Criteria Update Expert Panel, which this measure is based on, included review of 60 systematic reviews and meta analyses, 49 randomized control trials (RTCs) and 233 observational studies and other types of publications.

1c.6 Quality of Body of Evidence (*Summarize the certainty or confidence in the estimates of benefits and harms to patients across studies in the body of evidence resulting from study factors. Please address: a) study design/flaws; b) directness/indirectness of the evidence to this measure (e.g., interventions, comparisons, outcomes assessed, population included in the evidence); and c) imprecision/wide confidence intervals due to few patients or events):* Systematic synthesis of research and expert opinion = Low

Overall, the quality of the evidence for each of the medications included in the Beers Criteria recommendations is good. See table under section 1c.16 for the quality of evidence rating for the recommendation for each medication or medication class. The table also includes the AGS 2015 Beers Criteria Update Expert Panel rating for the strength of the evidence supporting each recommendation. Definitions of these ratings are listed in section 1c.21.

1c.7 Consistency of Results across Studies *(Summarize the consistency of the magnitude and direction of the effect):* The studies consistently mention similar drugs. Since the bodies of evidence all relate to the original Beers list, they maintain consistency in process.

See section 1c.16 for a table that contains the Beers Criteria recommendations for each drug and drug class that are included in the measure.

1c.8 Net Benefit *(Provide estimates of effect for benefit/outcome; identify harms addressed and estimates of effect; and net benefit - benefit over harms)***:**
Each updated study contributes to the strength of the measure by updating the medication lists.

See section 1c.16 for a table that contains the Beers Criteria recommendations for each drug and drug class that are included in the measure. Evidence tables containing summaries of each study supporting the recommendations can be found on the American Geriatrics Society's website: http://www.americangeriatrics.org/.

1c.9 Grading of Strength/Quality of the Body of Evidence. Has the body of evidence been graded? No
Yes

1c.10 If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: N/A
See 1c.20

1c.11 System Used for Grading the Body of Evidence: Other
See 1c.21

1c.12 If other, identify and describe the grading scale with definitions: N/A

1c.13 Grade Assigned to the Body of Evidence: N/A
See 1c.16

1c.14 Summary of Controversy/Contradictory Evidence: N/A

1c.15 Citations for Evidence other than Guidelines*(Guidelines addressed below)*:
Families USA, Cost Overdose: Growth in Drug Spending for the Elderly, 1992-2010. (Washington, DC. Families USA), July 2000, p. 2.

Budnitz, D et al. National surveillance of emergency department visits for outpatient adverse drug events. JAMA 2006;296:1858-1866.

Zhan, C, et al. Potentially inappropriate medication use in the community-dwelling elderly. JAMA 2001; 286(22):2823-2868

Beers, M.H. Explicit criteria for determining potentially inappropriate medication use by the elderly. Arch Intern Med 1997; 157:1531-1536.

Fick, DM, et al.Updating the Beers criteria for potentially inappropriate medication use in older adults. Arch Intern Med 2003; 163:2716-2724.

Curtis, LH, et al. Inappropriate Prescribing for Elderly Americans in a Large Outpatient Population Arch Intern Med 2004; 164:1621-1625.

Simon, SR, et al. Potentially Inappropriate Medication Use by Elderly Persons in U.S. Health Maintenance Organizations, 2000-200, Journal of the American Geriatrics Society, 2005, Volume 53, Issue 2, 227-232

Fu AZ, et al. Inappropriate Medication Use and Health Outcomes in the Elderly, Journal of the American Geriatrics Society 2004; Volume 52, Issue 11, 1934-9.

MacKinnon NJ, et al. Indicators of Preventable Drug-related Morbidity in Older Adults: Use Within a Managed Care Organization. J Managed Care Pharm. 2003; 9:134-41.

Kaufman MB, et al. Effect of Prescriber Education on the Use of Medications Contraindicated in Older Adults in a Managed Medicare Population. J Manag Care Pharm 2005 April/May; 11(3):211-219.

Rothberg MB, Perkow PS, Liu F, et al. Potentially inappropriate medication use in hospitalized elders. J Hosp Med. 2008;3:91-102.

Gray, PharmD, Gardner, MD. Adverse Drug Events in the Elderly: An Ongoing Problem, Journal of Managed Care Pharmacy Sep. 2009; Vol. 15, No. 7.

AHRQ, National Quality Measures Clearinghouse. www.qualitymeasures.ahrq.gov (Accessed Web page: October 12, 2009

1c.16 Quote verbatim, the specific guideline recommendation *(Including guideline # and/or page #)*:
N/A

Language in the table below is taken verbatim from Table 2 (pages 5-10) of the *American Geriatrics Society 2015 Updated Beers Criteria for Potentially Inappropriate Medication Use in Older Adults*. Evidence tables containing summaries of each study supporting the recommendations can be found on the American Geriatrics Society's website: http://www.americangeriatrics.org/.

Organ System, Therapeutic Category, Drugs	Rationale	Recommendation	Quality of Evidence	Strength of Recommendation
Anticholinergics: First-generation antihistamines (p. 5) Brompheniramine Carbinoxamine Chlorpheniramine Clemastine Cyproheptadine Dexbrompheniramine Dexchlorpheniramine Dimenhydrinate Diphenhydramine (oral) Doxylamine Hydroxyzine Meclizine Promethazine Triprolidine	Highly anticholinergic; clearance reduced with advanced age, and tolerance develops when used as hypnotic; risk of confusion, dry mouth, constipation, and other anticholinergic effects or toxicity Use of diphenhydramine in situations such as acute treatment of severe allergic reaction may be appropriate	Avoid	Moderate	Strong
Antiparkinsonian agents (p. 5) Benztropine (oral) Trihexyphenidyl	Not recommended for prevention of extrapyramidal symptoms with antipsychotics; more-effective agents available for treatment of Parkinson disease	Avoid	Moderate	Strong
Antispasmodics (p. 5) Atropine (excludes ophthalmic) Belladonna alkaloids Clidinium-Chlordiazepoxide Dicyclomine Hyoscyamine Propantheline Scopolamine	Highly anticholinergic, uncertain effectiveness	Avoid	Moderate	Strong
Antithrombotics (p. 5) Dipyridamole, oral short-acting	May cause orthostatic hypotension; more effective alternatives available;	Avoid	Moderate	Strong

(does not apply to the extended release combination with aspirin)	intravenous form acceptable for use in cardiac stress testing			
Antithrombotics (p. 5) Ticlopidine	Safer, effective alternatives available	Avoid	Moderate	Strong
Anti-infective (p. 5) Nitrofurantoin	Potential for pulmonary toxicity, hepatoxicity, and peripheral neuropathy, especially with long-term use; safer alternatives available	Avoid in individuals with creatinine clearance <30 mL/min or for long-term suppression of bacteria	Low	Strong
Central alpha blockers (p. 6) Guanabenz Guanfacine Methyldopa Reserpine (>0.1 mg/d)	High risk of adverse CNS effects; may cause bradycardia and orthostatic hypotension; not recommended as routine treatment for hypertension	Avoid	Low	Strong
Central alpha blockers (p. 6) Disopyramide	Disopyramide is a potent negative inotrope and therefore may induce heart failure in older adults; strongly anticholinergic; other antiarrhythmic drugs preferred	Avoid	Low	Strong
Central alpha blockers (p. 6) Digoxin	Use in atrial fibrillation: should not be used as a first-line agent in atrial fibrillation, because more effective alternatives exist and it may be associated with increased mortality Use in heart failure: questionable effects on risk of hospitalization and may be associated with increased mortality in older adults with heart failure; in heart failure, higher dosages not associated with additional benefit and may increase risk of toxicity Decreased renal clearance of digoxin may lead to increased risk of toxic effects; further dose reduction may be necessary in patients with Stage 4 or 5 chronic kidney disease	If used for atrial fibrillation or heart failure, avoid dosages >0.125 mg/d	Dosage >0.125 mg/d: Moderate	Dosage >0.125 mg/d: Strong
Central alpha blockers (p. 6) Nifedipine, immediate release	Potential for hypotension; risk of precipitating myocardial ischemia	Avoid	High	Strong
Central Nervous System (p. 7) Antidepressants, alone or in combination Amitriptyline Amoxapine Clomipramine Desipramine Doxepin >6 mg/d Imipramine Nortriptyline	Highly anticholinergic, sedating, and cause orthostatic hypotension; safety profile of low dose doxepin (≤6 mg/d) comparable with that of placebo	Avoid	High	Strong

Paroxetine Protriptyline Trimipramine				
Central Nervous System (p. 7) Barbiturates Amobarbital Butabarbital Butalbital Mephobarbital Pentobarbital Phenobarbital Secobarbital	High rate of physical dependence, tolerance to sleep benefits, greater risk of overdose at low dosages	Avoid	High	Strong
Central Nervous System (p. 8) Meprobamate	High rate of physical dependence; very sedating	Avoid	Moderate	Strong
Central Nervous System (p. 8) Nonbenzodiazepine, benzodiazepine receptor agonist hypnotics Eszopiclone Zolpidem Zaleplon	Benzodiazepine-receptor agonists have adverse events similar to those of benzodiazepines in older adults (e.g., delirium, falls, fractures); increased emergency department visits and hospitalizations; motor vehicle crashes; minimal improvement in sleep latency and duration	Avoid	Moderate	Strong
Central Nervous System (p. 8) Ergoloid mesylates (dehydrogenated ergot alkaloids) Isoxsuprine	Lack of efficacy	Avoid	High	Strong
Endocrine (p. 8) Desiccated thyroid	Concerns about cardiac effects; safer alternatives available	Avoid	Low	Strong
Endocrine (p. 8) Estrogens with or without progestins	Evidence of carcinogenic potential (breast and endometrium); lack of cardioprotective effect and cognitive protection in older women. Evidence indicates that vaginal estrogens for the treatment of vaginal dryness are safe and effective; women with a history of breast cancer who do not respond to nonhormonal therapies are advised to discuss the risk and benefits of low-dose vaginal estrogen (dosages of estradiol <25 lg twice weekly) with their healthcare provider	Avoid oral and topical patch Vaginal cream or tablets: acceptable to use low-dose intravaginal estrogen for management of dyspareunia, lower urinary tract infections, and other vaginal symptoms	Oral and patch: High Vaginal cream or tablets: Moderate	Oral and patch: Strong Topical vaginal cream or tablets: Weak
Endocrine (p. 9) Megestrol	Minimal effect on weight; increases risk of thrombotic events and possibly death in older adults	Avoid	Moderate	Strong
Endocrine (p. 9) Sulfonylureas, long-duration Chlorpropamide Glyburide	Chlorpropamide: prolonged half-life in older adults; can cause prolonged hypoglycemia; causes syndrome of inappropriate antidiuretic hormone secretion Glyburide: higher risk of severe prolonged hypoglycemia in older adults	Avoid	High	Strong

Pain medications (p. 9) Meperidine	Not effective oral analgesic in dosages commonly used; may have higher risk of neurotoxicity, including delirium, than other opioids; safer alternatives available	Avoid, especially in individuals with chronic kidney disease	Moderate	Strong
Pain medications (p. 10) Non-cyclooxygenase-selective NSAIDs, oral: Indomethacin Ketorolac, includes parenteral	Indomethacin is more likely than other NSAIDs to have adverse CNS effects. Of all the NSAIDs, indomethacin has the most adverse effects. Increased risk of gastrointestinal bleeding, peptic ulcer disease, and acute kidney injury in older adults	Avoid	Moderate	Strong
Pain medications (p. 10) Pentazocine	Opioid analgesic that causes CNS adverse effects, including confusion and hallucinations, more commonly than other opioid analgesic drugs; is also a mixed agonist and antagonist; safer alternatives available	Avoid	Low	Strong
Pain medications (p. 10) Skeletal muscle relaxants Carisoprodol Chlorzoxazone Cyclobenzaprine Metaxalone Methocarbamol Orphenadrine	Most muscle relaxants poorly tolerated by older adults because some have anticholinergic adverse effects, sedation, increased risk of fractures; effectiveness at dosages tolerated by older adults questionable	Avoid	Moderate	Strong

1c.17 Clinical Practice Guideline Citation: N/A
American Geriatrics Society 2015 Beers Criteria Update Expert Panel. 2015. American Geriatrics Society 2015 Updated Beers Criteria for Potentially Inappropriate Medication Use in Older Adults. Journal of the American Geriatrics Society, 63(11): 2227-2246.

1c.18 National Guideline Clearinghouse or other URL: N/A
http://geriatricscareonline.org/ProductAbstract/american-geriatrics-society-updated-beers-criteria-for-potentially-inappropriate-medication-use-in-older-adults/CL001

1c.19 Grading of Strength of Guideline Recommendation. Has the recommendation been graded? No
Yes

1c.20 If guideline recommendation graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias:
The American Geriatrics Society 2015 Beers Criteria Update Expert Panel graded the evidence. The panel had expertise in geriatric medicine, nursing, pharmacy practice, research, and quality measures. Other factors that influenced selection of panel members were the desire to have interdisciplinary representation, a range of medical expertise, and representation from different practice settings (e.g., long-term care, ambulatory care, geriatric mental health, palliative care and hospice). In addition to the 13-member panel, representatives from the Centers for Medicare and Medicaid Services, National Committee for Quality Assurance, and Pharmacy Quality Alliance served as ex-officio members of the panel. Each expert panel member completed a disclosure form at the beginning of the guideline process that was shared with the entire panel at the start of each panel meeting and call. Panel members who disclosed affiliations or financial interests with commercial entities are listed in the disclosures section of the *American Geriatrics Society 2015 Updated Beers Criteria for Potentially Inappropriate Medication Use in Older Adults* article. Panel members were then asked to recuse themselves from discussions if they had a potential conflict of interest.

1c.21 System Used for Grading the Strength of Guideline Recommendation: Other

The American Geriatrics Society 2015 Beers Criteria Update Expert Panel used the Grades of Recommendation Assessment, Development, and Evaluation (GRADE) rating process to rate the quality of evidence. Each panelist independently rated the quality of evidence and strength of recommendation for each criterion using the American College of Physicians' Guideline Grading System (Qaseem et al., 2010), which is based on the GRADE scheme (The GRADE Working Group). The chart below is excerpted from the Beers Criteria article and contains the definitions for the quality of evidence ratings and the strength of recommendations.

Quality of Evidence	
High	Evidence includes consistent results from well-designed, well-conducted studies in representative populations that directly assess effects on health outcomes (≥2 consistent, higher-quality randomized controlled trials or multiple, consistent observational studies with no significant methodological flaws showing large effects)
Moderate	Evidence is sufficient to determine risks of adverse outcomes, but the number, quality, size, or consistency of included studies; generalizability to routine practice; or indirect nature of the evidence on health outcomes (≥1 higher-quality trial with >100 participants; ≥2 higher-quality trials with some inconsistency; ≥2 consistent, lower-quality trials; or multiple, consistent observational studies with no significant methodological flaws showing at least moderate effects) limits the strength of the evidence
Low	Evidence is insufficient to assess harms or risks in health outcomes because of limited number or power of studies, large and unexplained inconsistency between higher-quality studies, important flaws in study design or conduct, gaps in the chain of evidence, or lack of information on important health outcomes
Strength of Recommendation	
Strong	Benefits clearly outweigh harms, adverse events, and risks, or harms, adverse events, and risks clearly outweigh benefits
Weak	Benefits may not outweigh harms, adverse events, and risks
Insufficient	Evidence inadequate to determine net harms, adverse events, and risks

References:
Qaseem A, Snow V, Owens DK et al. The development of clinical practice guidelines and guidance statements of the American College of Physicians: Summary of methods. Ann Intern Med 2010;153:194–199.

The GRADE working group. GRADE guidelines—best practices using the GRADE framework. Journal of Clinical Epidemiology [on-line]. Available at http://www.gradeworkinggroup.org/publications/jce_series.htm

1c.22 If other, identify and describe the grading scale with definitions: N/A

1c.23 Grade Assigned to the Recommendation: N/A
See table under 1c.16 for each recommendation with assigned grade.

1c.24 Rationale for Using this Guideline Over Others: N/A

Based on the NQF descriptions for rating the evidence, what was the <u>developer's assessment</u> of the quantity, quality, and consistency of the body of evidence?
1c.25 Quantity: High **1c.26** Quality: High**1c.27** Consistency**:** High

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

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

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., the benefits or improvements in quality envisioned by use of this measure)*
Lowering the use of high-risk medications in the elderly population should decrease morbidity and mortality associated with adverse drug reactions.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
The following data are extracted from HEDIS data collection for Medicare Advantage Health Plans (including all HMO and PPO plans). Performance data is summarized at the health plan level and summarized by mean, standard deviation, and performance at the 10th, 25th, 50th, 75th and 90th percentile. Data is stratified by year.

While the average performance for the first rate (at least one high-risk medication) has decreased from 21.0% in 2012 to 13.2% in 2014, the large difference in performance between plans at the 90th and the 10th percentiles represents a gap in care. The performance for the second rate (dispensing of two different high-risk medications) has also shown steady improvement from 2012 to 2014. The average performance in 2014 was 2.1%, with not much variation in the rate between plans at the 90th and 10th percentiles. For this reason, the specification was revised to assess multiple prescribing events for the same high-risk medication. We expect that future performance on this new rate will be higher (i.e., worse) and more variable than the previously specified rate.

At least 1 high-risk medication

YEAR	N	MEAN	ST DEV	10TH (Better)	25TH	50TH	75TH	90TH (Worse)	Interquartile Range
2012	498	21.0%	6.4%	14.0%	16.5%	19.9%	24.5%	30.0%	8%
2013	494	18.0%	6.1%	11.5%	13.8%	16.7%	21.1%	25.8%	7.4%
2014*	488	13.2%	6.0%	7.6%	9.2%	11.6%	16.1%	21.7%	6.9%

*For 2014 the average eligible population was 22,043, with a standard deviation of 45,532

At least 2 different high-risk medications^

YEAR	N	MEAN	ST DEV	10TH (Better)	25TH	50TH	75TH	90TH (Worse)	Interquartile Range
2012	498	6.5%	2.9%	3.5%	4.7%	6.0%	7.8%	10.1%	3.1%
2013	494	3.1%	2.3%	1.1%	1.7%	2.4%	4.0%	6.0%	2.3%
2014*	488	2.1%	2.0%	0.6%	0.9%	1.4%	2.5%	4.6%	1.6%

*For 2014 the average eligible population was 22,043, with a standard deviation of 45,532
^Note: These results are based on a previous specification of the HEDIS measure in which the numerator was based on multiple prescribing events of different high-risk medications instead of the current specification which looks at multiple prescribing events for the same high-risk medication.

The data referenced are extracted from HEDIS data collection reflecting the most recent years of measurement for this measure. In 2014, HEDIS measures covered more than 171 million people from 814 HMOs and 353 PPOs. Below is a description of the denominator for this measure. It includes the number of health plans included in HEDIS data collection and the mean eligible population for the measure across health plans.

YEAR | N Plans | Mean Denominator Size per plan
2012 | 498 | 18,090
2013 | 494 | 19,833
2014 | 488 | 22,043

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*
NCQA does not currently collect performance data stratified by race, ethnicity, or language. Escarce et al. have described in detail the difficulty of collecting valid data on race, ethnicity and language at the health plan level (Escarce, 2011). While not specified in the measure, this measure can also be stratified by demographic variables, such as race/ethnicity or socioeconomic status, in order to assess the presence of health care disparities. The HEDIS Health Plan Measure Set contains two measures that can assist with stratification to assess health care disparities. The Race/Ethnicity Diversity of Membership and the Language Diversity of Membership measures were designed to promote standardized methods for collecting these data and follow Office of Management and Budget and Institute of Medicine guidelines for collecting and categorizing race/ethnicity and language data. In addition, NCQA's Multicultural Health Care Distinction Program outlines standards for collecting, storing and using race/ethnicity and language data to assess health care disparities. Based on extensive work by NCQA to understand how to promote culturally and linguistically appropriate services among plans and providers, we have many examples of how health plans have used HEDIS measures to design quality improvement programs to decrease disparities in care.
Escare J.J., Carreon R., Vesolovskiy G., and Lawson E.H. 2011. Collection Of Race And Ethnicity Data By Health Plans Has Grown Substantially, But Opportunities Remain To Expand Efforts. Health Affairs 20(10): 1984-1991.
Centers for Disease Control and Prevention. 2010. Vital Signs. http://www.cdc.gov/VitalSigns/pdf/2010-07-vitalsigns.pdf (Accessed July 8, 2011).

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
While disparities for this measure have not been well studied, there is some evidence to suggest that women are more likely to receive a potentially inappropriate medication than men. A retrospective cohort study of 966,000 men and women treated by the Veteran's Health Administration showed that women were more likely than men to receive medications that may have harmful interactions with chronic conditions as described by the Beers Criteria (Bierman et al., 2007).
Bierman, A.S., M.J.V. Pugh, I. Dhalla, M. Amuan, B.G. Fincke, A. Rosen, D.R. Berlowitz. 2007. "Sex differences in inappropriate prescribing among elderly veterans." The American Journal of Geriatric Pharmacotherapy, 5(2):147-161.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, A leading cause of morbidity/mortality, High resource use, Patient/societal consequences of poor quality, Severity of illness
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
There is clinical consensus that in the elderly certain medications are associated with increased risk of harm from drug side-effects and drug toxicity; these medications pose a concern for patient safety. Use of potentially inappropriate medications (PIMs) in the elderly can lead to poor health outcomes including adverse drug events, confusion, falls, hospitalizations and even death. Despite

widely-accepted medical consensus that certain drugs increase the risk of harm to the elderly and should generally be avoided, these drugs are still frequently prescribed to the elderly. In a study of health outcomes, 40% of individuals 65 and older filled at least 1 potentially inappropriate medication (PIM) and 13% filled 2 or more (Fick et al. 2008). In this population, 14.3% of those who had at least one PIM had a drug-related problem, whereas only 4.7% of those with no PIMs had a drug-related problem. PIM use in the elderly has been connected to increased hospitalization and increased risk of death (Lau et al., 2004). Preventing poor health effects from use of PIMs is expected to be a growing concern with the increasing population of adults over 65, longer life expectancies and the introduction of new medications (Rothberg et al., 2008).

Reducing use of PIMs in the elderly also represents an opportunity to reduce the costs associated with harm from medications (e.g., hospitalizations from drug toxicity) and encourage clinicians to consider alternative, safer medications. Conservative estimates of extra costs due to potentially inappropriate medications in the elderly average $7.2 billion a year (Fu, 2007). The annual direct costs of preventable ADEs in the Medicare population have been estimated to exceed $800 million (Institute of Medicine, 2007). Reducing unnecessary prescribing will also help to reduce cost, given that the elderly population represent one third of all prescription drug expenditures in the U.S. but comprises only 13 percent of the population (Families USA, 2000). While expenditures for prescription drugs in the US are disproportionately clustered among those 65 years and older, this population is twice as likely as those below age 65 to experience adverse drug events and is almost seven times as likely to be hospitalized for adverse drug events (Budnitz, 2006).

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

Budnitz, D., D.A. Pollock, K.N. Widenbach, A.B. Mendelson, T.J. Schroeder, and J.L. Annest. 2006. "National Surveillance of Emergency Department Visits for Outpatient Adverse Drug Events." Journal of the American Medical Association 296:1858-1866.

Families USA, Cost Overdose: Growth in Drug Spending for the Elderly, 1992-2010. 2000. Washington, DC: Families USA. July, p. 2.

Fick, D.M., L.C. Mion, M.H. Beers, J.L. Waller. 2008. "Health Outcomes Associated with Potentially Inappropriate Medication Use in Older Adults." Research in Nursing & Health. 31(1): 42-51.

Fick, D.M., and T.P. Selma. 2012. 2012 American Geriatrics Society Beers Criteria: New Year, New Criteria, New Perspective. The American Geriatrics Society.

Fu, A.Z., J.Z. Jiang, J.H. Reeves, J.E. Funcham, G.G. Liu, M. Perri. 2007. "Potentially Inappropriate Medication Use and Healthcare Expenditures in the US Community-Dwelling Elderly." Medical Care 45: 472-6.

Institute of Medicine (IOM). 2007. Preventing Medication Errors/Committee on Identifying and Preventing Medication Errors. Ed. Aspden P., J.A. Wolcott, J.L. Bootman, L.R. Cronenwatt LR. Quality Chasm Series. Washington, DC: National Academy Press.
Lau, D.T.. J.D. Kasper, D.E. Potter, A. Lyles. 2004 "Potentially Inappropriate Medication Prescriptions Among Elderly Nursing Home Residents: Their Scope and Associated Resident and Facility Characteristics." Health Services Research 39(5): 1257-1276.
Rothberg, M.B., P.S. Perkow, F. Liu, B. Korc-Grodzicki, M.J. Brennan, S. Bellantonio, M. Heelon, P.K. Lindenauer. 2008. "Potentially Inappropriate Medication Use in Hospitalized Elders." Journal of Hospital Medicine. 3: 91-102.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

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 subcriteria 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):*

Prevention

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Safety : Medication Safety

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

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

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)*
No data dictionary **Attachment:**

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Since the last endorsement, the list of medications used in this measure has been updated to reflect the most current recommendations included in the American Geriatrics Society 2015 Updated Beers Criteria for Potentially Inappropriate Medication Use in Older Adults. The second indicator in the measure has also been changed to assess multiple prescribing events of the same high-risk medication instead of the prescribing of multiple different high-risk medications.

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)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Numerator 1: Patients who received at least one high-risk medication during the measurement year.

Numerator 2: Patients who received at least two prescriptions for the same high-risk medication during the measurement year.

For both numerators a lower rate indicates better performance.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
The measurement year (12-month period).

S.6. **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, 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.
Patients who had at least one dispensing event for a high-risk medication during the measurement year. Follow the steps below to identify numerator compliance. Include patients who meet criteria in more than one step only once in the numerator. Do not include denied claims.

Step 1: Identify patients with at least one dispensing event (any days supply) during the measurement year for a medication in Table DAE-A. These patients are compliant for Numerator 1.

Step 2: Identify patients with a single dispensing event during the measurement year for a medication in Table DAE-B where days supply exceeds the days supply criteria listed for the medication. These patients are compliant for Numerator 1. For medications dispensed during the measurement year, sum the days supply and include any days supply that extends beyond December 31 of the measurement year. For example, a prescription of a 90-days supply dispensed on December 1 of the measurement year counts as a 90-days supply.

Step 3: Identify patients with a single dispensing event during the measurement year for a medication in Table DAE-C where average daily dose exceeds the average daily dose criteria listed for the medication. These patients are compliant for Numerator 1. To calculate average daily dose multiply the quantity of pills dispensed by the dose of each pill and divide by the days supply. For example, a prescription for a 30-days supply of digoxin containing 15 pills, .250 mg each pill, has an average daily dose of 0.125 mg. To calculate average daily dose for elixirs and concentrates, multiply the volume dispensed by daily dose and divide by the days supply. Do not round when calculating average daily dose.

Numerator 2:
Patients who had at least two dispensing events for the same high-risk medication during the measurement year.
Follow the steps below to identify numerator compliance. Include patients who meet criteria in more than one step only once in the numerator. Do not include denied claims.

Step 1: Identify patients with two or more dispensing events (any days supply) on different dates of service during the measurement year for a medication in Table DAE-A. The dispensing events must be for the same drug as identified by the Drug ID in the NDC list. These patients are compliant for Numerator 2.

Step 2: For each patients identify all dispensing events during the measurement year for medications in Table DAE-B. Identify patients with two or more dispensing events on different dates of service for medications in the same medication class (as identified in the Description column). For example, a prescription for zolpidem and a prescription for zaleplon are considered two dispensing events for medications in the same medication class (these drugs share the same description: Nonbenzodiazepine hypnotics). Sum the days supply for prescriptions in the same medication class. Identify patients with two or more dispensing events for medications of the same medication class where the summed days supply exceeds the days supply criteria listed for the medication. These patients are compliant for Numerator 2. For medications dispensed during the measurement year sum the days supply and include any days supply that extends beyond December 31 of the measurement year. For example, a prescription of a 90-days supply dispensed on December 1 of the measurement year counts as a 90-days supply.
- Note: The intent is to identify all patients who had multiple dispensing events where the summed days supply exceeds the days supply criteria; there is no requirement that each dispensing event exceed the days supply criteria.

Step 3: For each patient identify all dispensing events during the measurement year for medications in Table DAE-C where average daily dose exceeds the average daily dose criteria listed for the medication. Identify patients with two or more dispensing events on the same or different dates of service that exceed the average daily dose criteria for the same drug as identified by the Drug ID in the NDC list (do not include drugs with a single dispensing event). These patients are compliant for Numerator 2. To calculate average daily dose for each dispensing event, multiply the quantity of pills dispensed by the dose of each pill and divide by the days supply. For example, a prescription for a 30-days supply of digoxin containing 15 pills, .250 mg each pill, has an average daily dose of 0.125 mg. To calculate average daily dose for elixirs and concentrates, multiply the volume dispensed by daily dose and divide by the days supply. Do not round when calculating average daily dose.

HIGH-RISK MEDICATIONS (Table DAE-A)
Anticholinergics, First-generation antihistamines:
Brompheniramine, Carbinoxamine, Chlorpheniramine, Clemastine, Cyproheptadine, Dexbrompheniramine, Dexchlorpheniramine, Diphenhydramine (oral), Dimenhydrinate, Doxylamine, Hydroxyzine, Meclizine, Promethazine, Triprolidine

Anticholinergics, anti-Parkinson agents:
Benztropine (oral), Trihexyphenidyl

Antispasmodics:
Atropine (exclude ophthalmic), Bellandonna alkaloids, Clidinium-Chlordiazepoxide, Dicyclomine, Hyoscyamine, Propantheline, Scopolamine

Antithrombotics:
Dipyridamole, oral short-acting (does not apply to the extended-release combination with aspirin), Ticlopidine

Cardiovascular, alpha agonists, central:
Guanabenz, Guanfacine, Methyldopa

Cardiovascular, other:
Disopyramide, Nifedipine (immediate release)

Central nervous system, antidepressants:
Amitriptyline, Clomipramine, Imipramine, Trimipramine, Amoxapine, Desipramine, Nortiptyline, Paroxetine, Protriptyline

Central nervous system, barbiturates:
Amobarbital, Butabarbital, Butalbital, Mephobarbital, Pentobarbital, Phenobarbital, Secobarbital

Central nervous system, vasodilators:
Ergot mesylates, Isoxsuprine

Central nervous system, other:
Meprobamate

Endocrine system, estrogens with or without progestins; include only oral and topical patch products:
Conjugated estrogen, Esterified estrogen, Estradiol, Estropipate

Endocrine system, sulfonylureas, long-duration:
Chlorpropamide, Glyburide

Endocrine system, other:
Desiccated thyroid, Megestrol

Pain medications, skeletal muscle relaxants:
Carisoprodol, Chlorzoxazone, Cyclobenzaprine, Metaxalone, Methocarbamol, Orphenadrine

Pain medications, other:
Indomethacin, Ketorolac (includes parenteral), Meperidine, Pentazocine

---
HIGH-RISK MEDICATIONS WITH DAYS SUPPLY CRITERIA (Table DAE-B)

Anti-infectives, other (greater than 90 days supply, days supply criteria):
Nitrofurantoin, Nitrofurantoin macrocrystals, Nitrofurantoin macrocrystals-monohydrate

Nonbenzodiazepine hypnotics (greater than 90 days supply, days supply criteria):
Eszopiclone, Zolpidem, Zaleplon

HIGH-RISK MEDICATIONS WITH AVERAGE DAILY DOSE CRITERIA (Table DAE-C)
Alpha agonists, central (greater than 0.1 mg/day, average daily dose criteria):
Reserpine

Cardiovascular, other (greater than 0.125 mg/day, average daily dose criteria):
Digoxin

Tertiary TCAs (as single agent or as part of combination products), (greater than 6 mg/day, average daily dose criteria):
Doxepin

Note: NCQA will post a comprehensive list of medications and NDC codes to www.ncqa.org by November 2016. For medications in Table DAE-A and DAE-C, identify different drugs using the Drug ID field located in the NDC list on NCQA's Web site (www.ncqa.org), posted by November, 2016.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
All patients 65 years of age and older.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, Populations at Risk : Dual eligible beneficiaries, Populations at Risk : Individuals with multiple chronic conditions, Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
All patients that are 66 years of age and older as of December 31 of the measurement year.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Patients who were enrolled in hospice care at any time during the measurement year.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
N/A

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
N/A

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
N/A

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
N/A

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **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.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
Step 1. Determine the denominator: All patients 66 years of age and older as of the end (e.g., December 31) of the measurement year.

Step 2: Identify numerator 1: Individuals in the denominator who have received at least one high-risk medication (see definition of

high-risk medications for numerator 1 in section S.6) during the measurement year.

Step 3: Identify numerator 2: Individuals in the denominator who have received at least two prescriptions for the same high-risk medication (see definition of high-risk medications for numerator 2 in section S.6) during the measurement year.

Step 4: Calculate the rates: Rate 1: Numerator 1 divided by the denominator; Rate 2: Numerator 2 divided by the denominator.

Note: for this measure a lower rate indicates better performance.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
N/A

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
N/A

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims, Electronic Clinical Data, Electronic Clinical Data : Pharmacy

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
This measure is based on administrative claims collected in the course of providing care to health plan members. NCQA collects the Healthcare Effectiveness Data and Information Set (HEDIS) data for this measure directly from Health Management Organizations and Preferred Provider Organizations via NCQA's online data submission system.

S.25. **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.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Health Plan, Integrated Delivery System

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic, Pharmacy
If other:

S.28. **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

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
DAE_Testing_Final.docx

NQF #: 0022 **NQF Project:** Patient Safety Measures-Complications Project, (Patient Safety Phase 3)

2. RELIABILITY & 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. **(evaluation criteria)**
Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate field. Supplemental materials may be referenced or attached in item 2.1. See guidance on measure testing.

2a2. Reliability Testing. (*Reliability testing was conducted with appropriate method, scope, and adequate demonstration of reliability.*)

2a2.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
HEDIS Health Plan performance data from 2010

HEDIS Health Plan performance data from 2012-2014

2a2.2 Analytic Method *(Describe method of reliability testing & rationale)*:
Reliability was estimated by using the beta-binomial model. Beta-binomial is a better fit when estimating the reliability of simple pass/fail rate measures as is the case with most HEDIS® health plan measures. The beta-binomial model assumes the plan score is a binomial random variable conditional on the plan´s true value that comes from the beta distribution. The beta distribution is usually defined by two parameters, alpha and beta. Alpha and beta can be thought of as intermediate calculations to get to the needed variance estimates. The beta distribution can be symmetric, skewed or even U-shaped.

Reliability used here is the ratio of signal to noise. The signal in this case is the proportion of the variability in measured performance that can be explained by real differences in performance. A reliability of zero implies that all the variability in a measure is attributable to measurement error. A reliability of one implies that all the variability is attributable to real differences in performance. The higher the reliability score, the greater is the confidence with which one can distinguish the performance of one plan from another. A reliability score greater than or equal to 0.7 is considered very good.

2a2.3 Testing Results *(Reliability statistics, assessment of adequacy in the context of norms for the test conducted)*:
Reliability for this measure was calculated as 0.99882 for one prescription and 0.99819 for two or more prescriptions.

Using 2014 HEDIS Health Plan performance data, reliability for this measure was calculated as 0.99814 for receipt of one or more high-risk prescriptions and 0.99594 for receipt of two or more high-risk prescriptions. Reliability scores can vary from 0.0 to 1.0. A score of zero implies that all variation is attributed to measurement error (noise) whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance (signal). Generally, a minimum reliability score of 0.7 is used to indicate sufficient signal strength to discriminate performance between accountable entities. The testing suggests that both indicators in this measure are highly reliable.

2b. VALIDITY. Validity, Testing, including all Threats to Validity: H☐ M☐ L☐ I ☐

2b1.1 Describe how the measure specifications *(measure focus, target population, and exclusions)* **are consistent with the evidence cited in support of the measure focus (***criterion 1c)* **and identify any differences from the evidence:**
The measure focuses on reducing risk of adverse drug events in the elderly population. The evidence is consistent with the focus and scope of this measure.

2b2. Validity Testing. (*Validity testing was conducted with appropriate method, scope, and adequate demonstration of validity.)*

2b2.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a*

sample, characteristics of the entities included):
The measure aligns with current evidence.

Validity statistics were calculated from 2014 HEDIS Health Plan performance data that included 488 Medicare health plans. This included all Medicare health plans submitting data to NCQA for HEDIS. The plans were geographically diverse and varied in size. The average (mean) eligible population for this measure across health plans was 22,043.

2b2.2 **Analytic Method** *(Describe method of validity testing and rationale; if face validity, describe systematic assessment)*:
NCQA tested the measure for face validity using a panel of stakeholders with specific expertise in measurement and women's health. This panel included representatives from key stakeholder groups geriatricians, health plans, Medicare officials and researchers. Experts reviewed the results of the field test and assessed whether the results were consistent with expectations, whether the measure represented quality care, and whether we were measuring the most important aspect of care in this area.

Method of Assessing Face Validity: This measure was tested for face validity with two panels of experts. See Additional Information: Ad.1. Workgroup/Expert Panel Involved in Measure Development for names and affiliations of expert panel members.

- The Geriatric Measurement Advisory Panel (GMAP) included 11 experts in geriatrics, including representation by consumers, health plans, health care providers and policy makers.
- NCQA's Committee on Performance Measurement (CPM) oversees the evolution of the measurement set and includes representation by purchasers, consumers, health plans, health care providers and policy makers. This panel is made up of 16 members. The CPM is organized and managed by NCQA and reports to the NCQA Board of Directors and is responsible for advising NCQA staff on the development and maintenance of performance measures. CPM members reflect the diversity of constituencies that performance measurement serves; some bring other perspectives and additional expertise in quality management and the science of measurement.

Method of Testing Construct Validity: We tested for construct validity by exploring whether the two rates within this measure were correlated with each other and with another measure of medication safety. We hypothesized that organizations that perform well on one of the indicators should perform well on the other indicator as well as the other medication safety measure. To test these correlations we used a Pearson correlation test. This test estimates the strength of the linear association between two continuous variables; the magnitude of correlation ranges from -1 and +1. A value of 1 indicates a perfect linear dependence in which increasing values on one variable is associated with increasing values of the second variable. A value of 0 indicates no linear association. A value of -1 indicates a perfect linear relationship in which increasing values of the first variable is associated with decreasing values of the second variable.

2b2.3 **Testing Results** *(Statistical results, assessment of adequacy in the context of norms for the test conducted; if face validity, describe results of systematic assessment)*:
This measure was deemed valid by the expert panel.

Results of Face Validity Assessment:
This measure was developed to address high-risk medication use in the elderly. NCQA and the GMAP worked together to assess which medications to include based on recommendations in the AGS Beers Criteria. The measure was field-tested from 2004-2005. After reviewing field test results the CPM recommended to send the measure to public comment with a majority vote in 2006. The measure was released for Public Comment in 2006 prior to publication in HEDIS. The CPM recommended moving this measure to first year data collection by a majority vote. The measure was then introduced in HEDIS 2007. Organizations reported the measures in the first year and the results were analyzed for public reporting in the following year. The CPM recommended moving this measure public reporting with a majority vote. In summary, the measure was deemed to have the desirable attributes of a HEDIS measure in 2006 (relevance, scientific soundness, and feasibility). These results indicate the MAPs and CPM showed agreement that the measures as specified will accurately differentiate quality across health plans. Our interpretation of these results is that this measure has sufficient face validity.

Results of Construct Validity Testing: The results in Table 1a indicate that there was a high correlation between the first and second rate in the measure. There were moderate correlations between both rates and the four rates in the other medication safety measure.

Table 1a. Correlations among both rates in the measure and a drug-disease interaction measure[1]

Measure	Pearson Correlation Coefficients

	Rate 1: One high-risk medication	Rate 2: Two high-risk medications
Rate 1: One high-risk medication		
Rate 2: Two high-risk medications	.8745	
Drug-disease interaction: History of Falls	0.307	.2735
Drug-disease interaction: Dementia	0.454	.4390
Drug-disease interaction: Chronic Kidney Disease	0.367	.3552
Drug-disease interaction: Total	0.386	.3913

Note: All correlations are significant at p<.05

[1]The *Potentially Harmful Drug-Disease Interactions in the Elderly* measure has four rates. The first rate assesses the percentage of patients 65 and older with a history of falls who received a high-risk medication. The second rate assesses the percentage of patients 65 and older with dementia who received a high-risk medication. The third rate assesses the percentage of patients 65 and older with chronic kidney disease who received a high-risk medication. The fourth rate is the sum of the three numerators divided by the sum of the three denominators for the three previous rates. Note: "high-risk" medications for each condition are based on recommendations in Table 3 of the American Geriatrics Society Beers Criteria.

Coefficients with absolute value of less than 0.3 are generally considered indicative of weak associations whereas absolute values of 0.3 or higher denote moderate to strong associations. The significance of a correlation coefficient is evaluated by testing the hypothesis that an observed coefficient calculated for the sample is different from zero. The resulting p-value indicates the probability of obtaining a difference at least as large as the one observed due to chance alone. We used a threshold of 0.05 to evaluate the test results. P-values less than this threshold imply that it is unlikely that a non-zero coefficient was observed due to chance alone. The results confirmed the hypothesis that rates in the measure are correlated with each other as well as with another measure of medication safety, suggesting they represent the same underlying quality construct of prescribing inappropriate medications for patients with the corresponding illnesses. These results indicate the measure is a valid measure of a plan's quality at managing use of high-risk medications in the elderly.

POTENTIAL THREATS TO VALIDITY. (*All potential threats to validity were appropriately tested with adequate results.*)

2b3. Measure Exclusions. (*Exclusions were supported by the clinical evidence in 1c or appropriately tested with results demonstrating the need to specify them.*)

2b3.1 **Data/Sample for analysis of exclusions** *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)***:**
NCQA currently allows health plans for optional exclusion to their results. NCQA does not conduct the annual analysis applied to a sample. In measure development, field testing and any re-analysis for update, we investigate and validate the effect reliability exclusion applied to the eligible denominator.

This measure has only one exclusion for individuals who are in hospice during the measurement year.

2b3.2 **Analytic Method** (*Describe type of analysis and rationale for examining exclusions, including exclusion related to patient preference)***:**
N/A

2b3.3 **Results** *(Provide statistical results for analysis of exclusions, e.g., frequency, variability, sensitivity analyses):*
N/A

2b4. Risk Adjustment Strategy. (*For outcome measures, adjustment for differences in case mix (severity) across measured entities was appropriately tested with adequate results.*)

2b4.1 **Data/Sample** *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)***:**
N/A

2b4.2 **Analytic Method (***Describe methods and rationale for development and testing of risk model or risk stratification including*

selection of factors/variables):
N/A

2b4.3 Testing Results *(Statistical risk model: Provide quantitative assessment of relative contribution of model risk factors; risk model performance metrics including cross-validation discrimination and calibration statistics, calibration curve and risk decile plot, and assessment of adequacy in the context of norms for risk models. Risk stratification: Provide quantitative assessment of relationship of risk factors to the outcome and differences in outcomes among the strata)*:
N/A

2b4.4 If outcome or resource use measure is not risk adjusted, provide rationale and analyses to justify lack of adjustment:
The measure assesses the use of high-risk medication in a general elderly population; risk adjustment is not indicated.

2b5. Identification of Meaningful Differences in Performance. (*The performance measure scores were appropriately analyzed and discriminated meaningful differences in quality.*)

2b5.1 Data/Sample *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
Data analysis demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful differences in performance.

2b5.2 Analytic Method *(Describe methods and rationale to identify statistically significant and practically/meaningfully differences in performance)*:
Comparison of means and percentiles; analysis of variance against established benchmarks: if sample size is >400, we would use an analysis of variance.

2b5.3 Results *(Provide measure performance results/scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)*:

One prescription	2009	2008	2007
N	294	278	244
MEAN	23	23.4	23.2
STDEV	8.99	9.03	9.26
STDERR	0.52	0.54	0.59
MIN	3.87	2.48	0.53
MAX	56.2	54.9	58.9
P10	13.3	12.8	12.1
P25	16.6	16.7	16.6
P50	22.2	22.9	22.8
P75	28.6	28.7	29.2
P90	35.2	33.9	34.8

At least 2 prescriptions	2009	2008	2007
N	294	278	244
MEAN	5.68	6.04	5.98
STDEV	4.23	4.38	4.45
STDERR	0.25	0.26	0.28
MIN	0.22	0	0
MAX	27.0	28.5	29.0
P10	1.6	1.72	1.8
P50	4.67	5.13	5.15
P75	7.69	7.79	8.25
P90	10.7	11.0	10.8

2012 to 2014 HEDIS Health Plan Performance Data
At least one high-risk prescription

	Number of Plans	Mean	Standard Deviation	Min	Max	10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile

2012	498	21.0	6.4	5.5	54.6	14.0	16.5	19.9	24.5	30.0
2013	494	18.0	6.1	1.0	50.5	11.5	13.8	16.7	21.1	25.8
2014	488	13.2	6.0	2.6	46.8	7.6	9.2	11.6	16.1	21.7

At least two high-risk prescriptions

	Number of Plans	Mean	Standard Deviation	Min	Max	10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile
2012	498	6.5	2.9	1.2	25.2	3.5	4.7	6.0	7.8	10.1
2013	494	3.1	2.3	0.0	20.6	1.1	1.7	2.4	4.0	6.0
2014	488	2.1	2.0	0.0	20.8	0.6	0.9	1.4	2.5	4.6

2b6. Comparability of Multiple Data Sources/Methods. (*If specified for more than one data source, the various approaches result in comparable scores.*)

The previous testing described in the sections below is not applicable for comparing multiple data sources as only one data source is used to calculate the measure (pharmacy claims).

2b6.1 **Data/Sample** *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
~~For the field test, NCQA required participating plans to provide data beyond what would normally be necessary to compute these measures. For purposes of the field test, the measurement year was 2002 and 2003. For each measure, the participating plans were asked to provide patient enrollment data and pharmacy data from administrative data systems for the entire measure eligible population.~~

~~NOTE: At the time of field testing, the measure was called "Drugs to be avoided in the elderly: a. Patients who receive at least one drug to be avoided, b. Patients who receive at least two different drugs to be avoided."~~

2b6.2 **Analytic Method** *(Describe methods and rationale for testing comparability of scores produced by the different data sources specified in the measure)*:
~~The purpose of field testing is to determine:~~
~~- The validity of the administrative algorithm to identify the target population (denominator) based upon the measurement period, continuous enrollment /exclusionary criteria~~
~~- The validity of administrative data to accurately capture medical processes delivered (i.e. tests) or diagnoses by comparing administrative results with data from a sample of medical records~~
~~- The feasibility of the measure specifications to identify the quality problem and to discriminate performance between health plans for the purposes of HEDIS public reporting.~~
~~- The reliability and feasibility of the measure specifications so that all health plans can capture the required data elements and can conduct programming~~

~~Based upon the field test results, NCQA made necessary revisions to the measure specifications so that it meets the Desirable Attributes of a HEDIS measure.~~

2b6.3 **Testing Results** *(Provide statistical results, e.g., correlation statistics, comparison of rankings; assessment of adequacy in the context of norms for the test conducted)*:
~~Percentage of members with at least one prescription:~~
~~Overall prescribing rates are high (92.6%) among elderly members (continuously enrolled for 12 months) with at least one prescription for any drug to be avoided based on Zahn or Beers list. However, this number may be inflated because the Beers list was more comprehensive and includes drugs considered to be "low severity" for potential patient safety, such as estrogen.~~

Drug "risk" categories:
On average, 45.2% of enrolled elderly are prescribed at least 1 of the drugs classified by Zahn in one of the 3 high-risk drug categories (plan range 18.7% - 86.7%).

Prescribing rates for drugs classified by Zahn into three high-risk drug categories are:
- About 7.7% of elderly enrollees get at least 1 "never appropriate" drug (plan range 2.4% to 9.1%)
- 27.5% get at least 1 "rarely appropriate" drugs (plan range 10.8% - 44.1%)
- 10% get at least 1 "sometimes indicated" drug (plan range 5.5% - 33.5%).

A more meaningful rate may be to show that on average 35.2% of members are prescribed at least 1 of the drugs from the "never appropriate" or "rarely appropriate" Zahn's high risk drug categories, (plan range 13.2% - 53.2%).

Another clinically meaningful quality indicator is to look at the percentage of members who received at least 2 prescriptions of different therapeutic classes for drugs to be avoided in the elderly. This represents a subset of members who are at increased risk of adverse drug events and patient safety from additional receipt of harmful drugs. Plan performance on this rate was 6%, plan range 1.1% - 9.3%. Women were more likely than men to receive 2 or more drugs (6.9% vs 4.2%), and older elderly patients ages 85 and older were less likely than younger elderly patients ages 65-74 years (4.2% vs 6.1%). These show specific areas for plans to target improvement.

When the Beers "high and low" severity drug risk categories are used, the extent of the quality problem appears to be much worse, although this classification includes a much broader group of drugs:
- Nearly three quarters (72.8%) of elderly enrollees get at least 1 prescription for a drug considered "high severity" and
- Over half (52.2%) of elderly enrollees get at least 1 prescription for a "low severity" drug.
- Less than 1% of elderly enrollees received a drug, Phenobarbital that is not classified by Beers, but considered "never appropriate" by Zahn. Note: Phenobarbital is also targeted by NCQA for a measure which requires annual drug level monitoring due to potential harms from drug toxicity.

On average, 74% of members are prescribed at least one of the drugs that are on the Beers list but not classified by Zahn (plan range 24.9% - 86.7%).

Number of prescriptions per member per year
NCQA calculated the average number of prescriptions for drugs to be avoided in elderly members per member per year, which included members who were not continuously enrolled for a full year. This method of reporting the information shows a wider range of performance between plans. This may be because some members receive multiple prescriptions for these drugs over the course of their membership. These figures are corroborated by the mean number of prescriptions for members who are enrolled for 12 months.

Based on the data, in 2002, each member could be receiving between 5 and 11 of the drugs on the combined Zahn and Beers list. This PMPY rate is calculated based on the number of drugs received by a member and the number of months he/she is enrolled and is reported as a number of prescriptions per year. This accounts for the time of potential exposure to the "risk" of receiving the drug while enrolled at the health plan. Are these actual cakculations…may want to put an transition sentence.

Using the Beers categories (high/low severity), on average a member receives:
- 3 - 6 "high severity" drug prescriptions per year
- 2 - 4.5 "low severity" drug prescriptions per year.

Using the Zhan category, on average a member receives:
- 0.9 – 3.6 prescriptions per year of any drugs in the 3 Zhan risk categories
- 0.1 - 0.3 prescriptions per year for "never appropriate" drugs
- 0.6 - 1.6 prescriptions per year for "rarely appropriate" drugs
- 0.5 - 1.7 prescriptions per year for "sometimes indicated" drugs,
- 3.8 – 7.0 prescriptions per year for drugs not classified by Zahn but which are "high" or "low" severity on the Beers list.
- 0.7 – 2.9 prescriptions per year for "never" or "rarely" appropriate drugs

2c. Disparities in Care: **H☐ M☐ L☐ I ☐ NA☐** (*If applicable, the measure specifications allow identification of disparities.*)

2c.1 If measure is stratified for disparities, provide stratified results *(Scores by stratified categories/cohorts)*: The measure is not

stratified to detect disparities. NCQA has participated with IOM and others in attempting to include information on disparities in measure data collection. However, at the present time, this data, at all levels (claims data, paper chart review, and electronic records), is not coded in a standard manner, and is incompletely captured. There are no consistent standards for what entity (physician, group, plan, employer) should capture and report this data. While "requiring" reporting of the data could push the field forward, it has been our position that doing so would create substantial burden with inability to use the data because of its inconsistency. At the present time, we agree with the IOM report that disparities are best considered by the use of zip code analysis which has limited applicability in most reporting situations. At the health plan level, for HEDIS health plan data collection, NCQA does have extensive data related to our use of stratification by insurance status (Medicare, Medicaid and private-commercial) and would strongly recommend this process where the data base supporting the measurement includes this information. However, we believe that the measure specifications should NOT require this since the measure is still useful where the data needed to determine disparities cannot be ascertained from the data available.

2c.2 **If disparities have been reported/identified (e.g., in 1b), but measure is not specified to detect disparities, please explain:**
N/A

2.1-2.3 **Supplemental Testing Methodology Information:**

Steering Committee: Overall, was the criterion, *Scientific Acceptability of Measure Properties*, met?
(*Reliability and Validity must be rated moderate or high*) Yes☐ No☐
Provide rationale based on specific subcriteria:

If the Committee votes No, STOP

3. Feasibility

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

3a. Byproduct of Care Processes

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

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score)
If other:

3b. Electronic Sources

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

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

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.**
No feasibility assessment **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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
NCQA recognizes that, despite the clear specifications defined for HEDIS measures, data collection and calculation methods may vary, and other errors may taint the results, diminishing the usefulness of HEDIS data for managed care organization (MCO) comparison. In order for HEDIS to reach its full potential, NCQA conducts an independent audit of all HEDIS collection and reporting processes, as well as an audit of the data which are manipulated by those processes, in order to verify that HEDIS specifications are met. NCQA has developed a precise, standardized methodology for verifying the integrity of HEDIS collection and calculation processes through a two-part program consisting of an overall information systems capabilities assessment followed by an evaluation of the MCO´s ability to comply with HEDIS specifications. NCQA-certified auditors using standard audit methodologies will help enable purchasers to make more reliable "apples-to-apples" comparisons between health plans.
The HEDIS Compliance Audit addresses the following functions:
1) information practices and control procedures
2) sampling methods and procedures
3) data integrity
4) compliance with HEDIS specifications
5) analytic file production
6) reporting and documentation

In addition to the HEDIS Audit, NCQA provides a system to allow “real-time” feedback from measure users. Our Policy Clarification Support System receives thousands of inquiries each year on over 100 measures. Through this system NCQA responds immediately to questions and identifies possible errors or inconsistencies in the implementation of the measure. This system is vital to the regular re-evaluation of NCQA measures.

Input from NCQA auditing and the Policy Clarification Support System informs the annual updating of all HEDIS measures including updating value sets and clarifying the specifications. Measures are re-evaluated on a periodic basis and when there is a significant change in evidence. During re-evaluation information from NCQA auditing and Policy Clarification Support System is used to inform evaluation of the scientific soundness and feasibility of the measure.

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)*.
Broad public use and dissemination of these measures is encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Use by health care physicians in connection with their own practices is not commercial use. Commercial use of a measure requires the prior written consent of NCQA. As used herein, “commercial use” refers to any sale, license or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.

4. Usability and Use

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

4a. Accountability and Transparency

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

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

Planned	Current Use (for current use provide URL)
	Public Reporting Health Plan Ratings http://www.ncqa.org/ReportCards/HealthPlans/HealthInsurancePlanRankings/HealthPlanRatingsPreview.aspx Annual State of Health Care Quality http://www.ncqa.org/tabid/836/Default.aspx CMS Physician Quality Reporting System (PQRS) http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/ CMS EHR INCENTIVE PROGRAM (MEANINGFUL USE) https://www.healthit.gov/providers-professionals/meaningful-use-definition-objectives (PQA Use of High-Risk Medications in the Elderly measure) CMS Medicare Part D https://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovContra/PartCDDataValidation.html Payment Program CMS Physician Quality Reporting System (PQRS) http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/ CMS EHR INCENTIVE PROGRAM (MEANINGFUL USE https://www.healthit.gov/providers-professionals/meaningful-use-definition-objectives Regulatory and Accreditation Programs HEDIS®-Health Plan http://www.ncqa.org/Programs/Accreditation/HealthPlanHP.aspx HEDIS®-ACO http://www.ncqa.org/Programs/Accreditation/AccountableCareOrganizationACO.aspx HEDIS®-Physician http://www.ncqa.org/Programs/Certification/PhysicianandHospitalQualityPHQ.aspx Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Annual State of Health Care Quality http://www.ncqa.org/tabid/836/Default.aspx

4a.1. For each CURRENT use, checked above, provide:

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

HEALTH PLAN RATINGS/REPORT CARDS: This measure is used to calculate health plan ratings which are reported in Consumer Reports and on the NCQA website. These ratings are based on performance on HEDIS measures among other factors. In 2012, a total of 455 Medicare Advantage health plans, 404 commercial health plans and 136 Medicaid health plans across 50 states were included in the ratings. In 2015 NCQA announced a change in methodology and changed Health Plan Rankings to Health Plan Ratings.
CMS EHR INCENTIVE PROGRAM (MEANINGFUL USE): The Medicare and Medicaid Electronic Health Care Record (EHR) Incentive Programs provide incentive payments to eligible professionals, eligible hospitals, and critical access hospitals (CAHs) as they adopt, implement, upgrade or demonstrate meaningful use of certified EHR technology.
CMS Medicare Part D: This measure is aligned with the Pharmacy Quality Alliance's Use of High-Risk Medications in the Elderly measure which is reported by Medicare Part D plans. Organizations contracted to offer Medicare Part D benefits are required to report data to CMS on a variety of measures. CMS has developed reporting standards and data validation specifications with respect to the Part D reporting requirements. These standards and specifications provide a review process for Medicare Advantage Organizations (MAOs), Cost Plans, and Part D sponsors to use to conduct data validation checks on their reported Part D data. The data validation is "retrospective," referring to the fact that it normally occurs in the year subsequent to the measurement year.

STATE OF HEALTH CARE ANNUAL REPORT: This measure is publically reported nationally and by geographic regions in the NCQA State of Health Care annual report. This annual report published by NCQA summarizes findings on quality of care. In 2012 the report included measures on 11.5 million Medicare Advantage beneficiaries in 455 Medicare Advantage health plans, 99.4 million members in 404 commercial health plans, and 14.3 million Medicaid beneficiaries in 136 plans across 50 states.
HEALTH PLAN ACCREDITATION: This measure is used in scoring for accreditation of Medicare Advantage Health Plans. In 2012, a total of 170 Medicare Advantage health plans were accredited using this measure among others covering 7.1 million Medicare beneficiaries. Health plans are scored based on performance compared to benchmarks.
HEDIS ACCOUNTABLE CARE ORGANIZATION ACCREDITATION: This measure is used in NCQA's ACO Accreditation program, that helps health care organizations demonstrate their ability to improve quality, reduce costs and coordinate patient care. ACO standards and guidelines incorporate whole-person care coordination throughout the health care system.
HEDIS PHYSICIAN ACCREDITATION: This measure is used in NCQA's Physician Accreditation program, that helps physicians demonstrate their ability to improve quality, reduce costs and coordinate patient care.

PHYSICIAN QUALITY REPORTING SYSTEM: This measure is used in the Physician Quality Reporting System (PQRS) which is a reporting program that uses a combination of incentive payments and payment adjustments to promote reporting of quality information by eligible professionals (EPs). Eligible professionals who satisfactorily report data on quality measures for covered Physician Fee Schedule services furnished to Medicare Part B beneficiaries (including Railroad Retirement Board and Medicare Secondary Payer) receive these payment incentives and adjustments.

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

4a.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.*)
N/A

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

Over the past three years the two rates in this measure have shown steady improvement across health plans (approximately 8% improvement over the past three years in the first rate and 4.4% in the second rate). See section 1b.2 for a summary of data from health plans. These data are nationally representative.

Performance rates among plans for the second rate assessing the receipt of two or more different high-risk medications have improved to a point where we see little room for continued improvement. Therefore, we've changed the rate to assess multiple dispensing events for the same high-risk medication, where we expect to see greater room for improvement.

Additionally, due to recent updates to the medications included in the measure, future rates may show greater room for improvement and variation in performance.

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
N/A

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
There were no identified unintended consequences for this measure during testing or since implementation. If this measure were to be implemented poorly, there is concern that it could lead to reduced access to medications. There will always be individual cases that will warrant the use of a potentially harmful medication and clinicians should weigh the risks and benefits of using these medications for their individual patients.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
Yes

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

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**
NCQA: Potentially Harmful Drug-Disease Interactions in the Elderly
This measure is being submitted as a new measure for NQF endorsement during this current Patient Safety project.

5a. Harmonization
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 completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
The Potentially Harmful Drug-Diseased Interactions in the Elderly (DDE) measure and NQF 0022 have a similar focus (measuring potentially inappropriate medication use in the elderly) and reporting level (health plan), however they have different target populations. The DDE measure targets patients with a specific condition or disease that can experience adverse effects when combined with certain medications that are recommended to be avoided for that condition. This measure (NQF 0022) targets a larger population of all older adults and assesses use of high-risk medications that have been recommended to be avoided in all older adults. The DDE measure is being submitted as a new measure for NQF endorsement during this current Patient Safety project.

5b. Competing Measures
The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

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

Appendix

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

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** National Committee for Quality Assurance
Co.2 **Point of Contact:** Bob, Rehm, nqf@ncqa.org, 202-955-1728-
Co.3 **Measure Developer if different from Measure Steward:** National Committee for Quality Assurance
Co.4 **Point of Contact:** Bob, Rehm, nqf@ncqa.org, 202-955-1728-

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.
Geriatric Measurement Advisory Panel (GMAP):
Wade Aubry, University of California, San Francisco
Arlene Bierman, Agency for Healthcare Research and Quality (AHRQ)
Patricia Bomba, Excellus BlueCross BlueSheild
Jennie Chin Hansen, American Geriatrics Society
Joyce Dubow, Consumer Advocate
Peter Hollmann, Brown University
Adrienne Mims, Alliant Quality
Steven Phillips, Sierra Health Services, Inc.
Eric G Tangalos, Mayo Clinic
Joan Weiss, Health Resources and Services Administration
Neil Wenger, UCLA Division of General Internal Medicine and RAND

Committee on Performance Measurement (CPM):
Bruce Bagley, MD, FAAFP, Senior Advisor to the Professional Satisfaction and Practice Sustainability effort at the American Medical Association
Andrew Baskin, MD, National Medical Director, Quality & Provider Performance Measurement, Aetna
Patrick Conway, MD, MSC, Chief Medical Officer and Deputy Administrator, Centers for Medicare and Medicaid Services
Jonathan D. Darer, MD, MPH, Chief Innovation Officer, Geisinger Health System
Helen Darling, Strategic Advisor, National Business Group of Health
Rebekah Gee, MD, MPH, FACOG, Assistant Professor, LSUHSC
Foster Gesten, MD, FACP, New York State Department of Health
Marge Ginsburg, Executive Director, Center for Healthcare Decisions
David Grossman, MD, MPH, Executive Medical Director, Population and Purchaser Strategy, Group Health
Christine S. Hunter, MD (Co- Chair), Chief Medical Officer, US Office of Personnel Management
Jeffery Kelman, MMSc, MD, Chief Medical Officer, United Stated Department of Health and Human Services
Bernadette Loftus, MD, Associate Executive Director for the Mid-Atlantic States, The Permanente Medical Group
J. Brent Pawlecki, MD, MMM, Chief Health Officer, The Goodyear Tire & Rubber Company
Susan Reinhard, PhD, RN, Senior Vice President, AARP Public Policy Institute
Eric C. Schneider, MD, MSc, FACP (Co-chair), Senior Vice President, Policy and Research, The Commonwealth Fund
Marcus Thygeson, MD, MPH, Chief Health Officer, Blue Shield of California

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2006
Ad.3 **Month and Year of most recent revision:** 05, 2016
Ad.4 **What is your frequency for review/update of this measure?** Approximately every 3 years, sooner if the clinical guidelines have changed significantly.
Ad.5 **When is the next scheduled review/update for this measure?** 12, 2017

Ad.6 **Copyright statement:** ©2006 by the National Committee for Quality Assurance
1100 13th Street, NW, Suite 1000
Washington, DC 20005
Ad.7 **Disclaimers:** These performance measures are not clinical guidelines and do not establish a standard of medical care, and have not been tested for all potential applications.
THE MEASURES AND SPECIFICATIONS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND.

Ad.8 **Additional Information/Comments:** NCQA Notice of Use. Broad public use and dissemination of these measures is encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Use by health care physicians in connection with their own practices is not commercial use. Commercial use of a measure requires the prior written consent of NCQA. As used herein, "commercial use" refers to any sale, license or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.

These performance measures were developed and are owned by NCQA. They are not clinical guidelines and do not establish a standard of medical care. NCQA makes no representations, warranties or endorsement about the quality of any organization or physician that uses or reports performance measures, and NCQA has no liability to anyone who relies on such measures. NCQA holds a copyright in these measures and can rescind or alter these measures at any time. Users of the measures shall not have the right to alter, enhance or otherwise modify the measures, and shall not disassemble, recompile or reverse engineer the source code or object code relating to the measures. Anyone desiring to use or reproduce the measures without modification for a noncommercial purpose may do so without obtaining approval from NCQA. All commercial uses must be approved by NCQA and are subject to a license at the discretion of NCQA. © 2012 by the National Committee for Quality Assurance

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 0450
De.2. **Measure Title:** Perioperative Pulmonary Embolism or Deep Vein Thrombosis Rate (PSI 12)
Co.1.1. **Measure Steward:** Agency for Healthcare Research and Quality
De.3. **Brief Description of Measure:** Perioperative pulmonary embolism or proximal deep vein thrombosis (secondary diagnosis) per 1,000 surgical discharges for patients ages 18 years and older. Excludes cases with principal diagnosis for pulmonary embolism or proximal deep vein thrombosis; cases with secondary diagnosis for pulmonary embolism or proximal deep vein thrombosis present on admission; cases in which interruption of vena cava occurs before or on the same day as the first operating room procedure; and obstetric discharges.
1b.1. **Developer Rationale:** Deep vein thrombosis (DVT) is the formation of a blood clot in a deep vein—usually in the leg or pelvic veins. The most serious complication of a proximal DVT is that the clot dislodges and can travel to the lungs, becoming a pulmonary embolus (PE). Venous thromboembolism (VTE) is common in the perioperative setting, especially after high-risk operations, and can be deadly. Clinical trials have demonstrated that mechanical and pharmacologic interventions can substantially reduce the risk of perioperative VTE among moderate and high-risk surgical patients, especially when these interventions are initiated before or immediately after surgery and continued until or after discharge. Case control studies have demonstrated that early ambulation after surgery can further reduce the risk of perioperative VTE among high-risk surgical patients who receive appropriate mechanical or pharmacologic prophylaxis. Effective and safe prophylactic measures are now available for most high risk patients, and numerous evidence-based guidelines have been published for the prevention of VTE (most notably by the American College of Chest Physicians and the American Academy of Orthopedic Surgeons).

As summarized in a 2015 AHRQ report on Preventing Hospital Associated Venous Thromboembolism (available at http://www.ahrq.gov/sites/default/files/wysiwyg/professionals/quality-patient-safety/patient-safety-resources/resources/vtguide/vteguide.pdf):
"Thromboprophylaxis for at-risk inpatients can reduce VTE by 30% to 65%, has a low incidence of major bleeding complications, and has well-documented cost-effectiveness… Numerous guidelines from authoritative bodies outlining appropriate use of thromboprophylaxis are available… yet study after study reflects unacceptably low rates of thromboprophylaxis in patients at risk… For example, a recent cross-sectional international study of almost 70,000 patients in 358 hospitals found that appropriate prophylaxis was administered in only 58.5% of surgical and 39.5% of medical inpatients at risk for VTE; another U.S. registry found only 42 percent of patients with hospital-associated DVT received prophylaxis within 30 days prior to diagnosis… This constellation of facts presents a powerful imperative for improvement."

"This "implementation gap" in VTE prophylaxis between evidence-based best practice and actual practice in the real world has not gone unnoticed as a major opportunity for improvement. In 2008, the U.S. Surgeon General produced a call-to-action document for VTE prevention… In addition, key goals for VTE prevention are in place from the National Quality Forum and the Joint Commission… VTE Prevention is one of the focus areas of the Partnership for Patients, a major effort from the Centers for Medicare & Medicaid Services (CMS) to foster accelerated improvement…. Reports commissioned by AHRQ called thromboprophylaxis the "number one" patient safety practice… and a 2013 update continues to list improved prophylaxis for VTE as a top 10 patient safety strategy to act on now…. The American Public Health Association has stated that the "disconnect between evidence and execution as it relates to DVT prevention amounts to a public health crisis…"

"Various strategies to improve the use of thromboprophylaxis have been demonstrated to be effective, including computerized order sets with electronic alerts, or pre-printed orders and quality improvement in the form of clinician education programs, audit, and feedback, but further efforts are required at improving the translation of data from clinical trials into clinical practice…" Use of PSI 12, and related measures developed by The Joint Commission, encourages providers to adopt the processes or structures of care of the best performing providers, and may empower consumers to select better performing providers or to adhere to recommended

prophylactic modalities.

S.4. **Numerator Statement:** Discharges, among cases meeting the inclusion and exclusion rules for the denominator, with a secondary ICD-9-CM or ICD-10-CM diagnosis code for proximal deep vein thrombosis or a secondary ICD-9-CM or ICD-10-CM diagnosis code for pulmonary embolism.
S.7. **Denominator Statement:** Surgical discharges, for patients ages 18 years and older, with any-listed ICD-9-CM or ICD-10-PCS procedure codes for an operating room procedure. Surgical discharges are defined by specific MS-DRG codes.
S.10. **Denominator Exclusions:** Exclude cases:

- with a principal ICD-9-CM or ICD-10-CM diagnosis code (or secondary diagnosis present on admission) for proximal deep vein thrombosis
- with a principal ICD-9-CM or ICD-10-CM diagnosis code (or secondary diagnosis present on admission) for pulmonary embolism
- where a procedure for interruption of vena cava occurs before or on the same day as the first operating room procedure*
- any-listed ICD-9-CM or ICD-10-PCS procedure code for extracorporeal membrane oxygenation (ECMO)
- any-listed ICD-9-CM or ICD-10-CM diagnosis code for acute brain or spinal injury present on admission
- MDC 14 (pregnancy, childbirth, and puerperium)
- with missing gender (SEX=missing), age (AGE=missing), quarter (DQTR=missing), year (YEAR=missing), or principal diagnosis (DX1=missing)

*If day of procedure is not available in the input data file, the rate may be slightly lower than if the information was available.

De.1. **Measure Type:** Outcome
S.23. **Data Source:** Administrative claims
S.26. **Level of Analysis:** Facility

IF Endorsement Maintenance – Original Endorsement Date: Jul 31, 2008 **Most Recent Endorsement Date:** Aug 09, 2012

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

Maintenance of Endorsement-- Preliminary Analysis

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

Criteria 1: Importance to Measure and Report

1a. Evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

Summary of evidence:

- The developers identified the evidence through formal environmental scans of the literature. Clinical trials have demonstrated that mechanical and pharmacologic interventions can substantially reduce the risk of perioperative VTE among moderate and high-risk surgical patients.
- Case control studies have demonstrated that early ambulation after surgery can further reduce the risk of perioperative VTE among high-risk surgical patients who receive appropriate mechanical or pharmacologic prophylaxis.
- Thromboprophylaxis for at-risk inpatients can reduce VTE by 30% to 65%, has a low incidence of major

bleeding complications, and has well-documented cost-effectiveness.

- Various strategies to improve the use of thromboprophylaxis have been demonstrated to be effective, including computerized order sets with electronic alerts, or pre-printed orders and quality improvement in the form of clinician education programs, audit, and feedback, but further efforts are required at improving the translation of data from clinical trials into clinical practice.

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*

Preliminary rating for evidence: ☒ Pass ☐ No Pass

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

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

- The developer provides a summary of performance data from 2011-2013 populated from the Healthcare Cost and Utilization Project database from a very large sample.
- The mean rate was 3.437 per 1000 surgical discharges in for 2011-2012 and 3.620 per 1000 surgical discharges in 2012-2013.

Disparities

- The developer provides rates stratified by gender, age, payer, race/ethnicity and residence. The rates vary by these characteristics but there is no indication of whether or not these differences are significant.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*

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)

1. Importance to Measure and Report

1a. Evidence to Support Measure Focus

Comments: ** This is an outcome measure that uses administrative claims data. Environmental scan of the literature shows that clinical trials have demonstrated that mechanical and pharma interventions can substantially reduce the risk of perioperative VTE among moderate ang high-risk surgical patients. Case control study results were also provided.

1b. Performance Gap

Comments: ** "HCUP performance data from 2011-2013 were provided; mean rate was 3.437/1000 surgical discharges in 2011-12 and 3.620 /1000 in 2012-13. Rates vary by gender, age, payer, race/ethnicity and residence but no indication of significance.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications
Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Administrative Claims

Specifications:

- This measure uses administrative claims to identify the rate of perioperative pulmonary embolism or proximal deep vein thrombosis (secondary diagnosis) per 1,000 surgical discharges for patients ages 18 years and older.
- The level of analysis for this measure is the hospital/acute care facility.
- The recommended time period for data is two years for users with a complete sample of hospital discharges; the developer notes that the signal variance parameters in the software assume at least a one-year time period, and that users may use longer time periods if desired.
- The denominator (surgical discharges) is defined using ICD-9/10 operating room procedure codes and Medicare Severity Diagnosis-Related Group (MS-DRG) codes.
- The numerator identifies cases with a secondary ICD-9-CM or ICD-10-CM diagnosis code for proximal deep vein thrombosis or a secondary ICD-9-CM or ICD-10-CM diagnosis code for pulmonary embolism.
- The measure is expressed as a risk-adjusted rate per 1,000 surgical discharges; the risk-adjusted rate is computed using indirect standardization as the observed rate divided by the expected rate, multiplied by the reference population rate.

Questions for the Committee :

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

SUMMARY OF TESTING

Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☐ **Yes** ☐ **No**

Method(s) of reliability testing

- To demonstrate reliability, the developer conducted a signal-to-noise analysis of the measure score.
- The developer assessed the measure's signal-to-noise ratio by comparing the degree to which risk adjusted rates differ across hospitals (the signal) to the degree of precision of the rates within hospitals (the noise):
  - The signal-to-noise ratio was calculated at the hospital level and then summarized across the entire population of US hospitals.
- The developer notes that hospital size has an impact on reliability, and that smaller hospitals have less reliable rates due to very small denominators (the number of patients at risk). For this reason, the overall signal-to-noise ratio for the measure is calculated as a weighted estimate, using a method that reduces the influence of smaller hospitals.

Results of reliability testing

- To report the results of reliability testing, the developer grouped hospitals into deciles by size, and provided the average signal-to-noise ratio for each decile, as well as an overall reliability score:

Hospital Size Decile	Number of Hospitals	Avg. Number of Discharges per Hospital in Decile	Avg. Signal-to-Noise Ratio for Hospitals in Decile
1 (smallest)	357	45.3	0.0561
2	358	228.5	0.1595
3	357	541.9	0.2895

4	358	998.3	0.4197
5	357	1,608.1	0.5362
6	358	2,372.7	0.6275
7	358	3,302.9	0.7024
8	357	4,548.3	0.7675
9	358	6,561.0	0.8298
10 (largest)	357	11,648.9	0.8975
Overall	3,575	3,184.9	0.7359

- The developer observes that signal-to-noise ratios were smaller for hospitals with fewer than approximately 100 qualifying discharges per year (average signal-to-noise ratio less than 0.42). For this reason, the developer recommends the use of 'smoothed rates', which bring scores toward the mean, particularly for smaller hospitals.
    - The developer notes that smoothed rates are implemented in the AHRQ software.
- The developer argues that there is no universally accepted threshold of "adequate" signal to noise ratio, stating that:
    - Different methods of calculating reliability and signal-to-noise (e.g., split sample or test-retest reliability of the data, different methods of calculating the hospital signal-to-noise ratio) result in different distributions of reliability scores. In addition, "adequate" depends on the specific application and judgment of the user. For instance, if a complication such as mortality is very important (e.g. leads to great harm to the patient) a lower reliability may be acceptable. However, the AHRQ QI program generally considers ratios between 0.4 – 0.8 as acceptable. It is rare to achieve reliability above 0.8, using hospital signal-to-noise ratios as an indicator of reliability.
- The developer considers this indicator to have a good overall signal-to-noise ratio at 0.74.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *What do you think about the developer's findings on the reliability of smaller hospitals and the developer's approach to addressing these issues?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

Guidance from the Reliability Algorithm :

[Box 1] Specifications precise and unambiguous → [Box 2] Empirical testing conducted on the measure as specified → [Box 4] Testing conducted at the measure score level → [Box 5] → Testing method described and appropriate → [Box 6] Moderate certainty or confidence that measure scores are reliable → [Box 6b]

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

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

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.

SUMMARY OF TESTING

Validity testing level ☐ **Measure score** ☐ **Data element testing against a gold standard** ☒ **Both**

Method of validity testing of the measure score:

- ☐ **Face validity only**
- ☒ **Empirical validity testing of the measure score**

Validity testing method:

- To demonstrate validity, the developer takes several approaches:
  - Data element validity
    - The developer notes that several studies have evaluated the validity of ICD coding (specifically around identification of postoperative PE/DVT) using medical record abstraction by trained nurses as a gold standard; the developer summarizes a number of such studies published in the peer review literature.
  - Measure score validity
    - To assess validity of the measure score, the developer focused on a specific critique of PSI 12: that the indicator should include only elective admissions, which typically have the index procedures on day 0 or 1 of the hospitalization.
      - To test this critique, the developer stratified PSI 12 events in the 2012 AHRQ QI Reference Population by preoperative length of stay (days before the index procedure) and calculated the PSI 12 rate for each stratum.
    - In addition, the developer utilized a structured panel review to evaluate face validity of the measure.
      - 7 members of a multispecialty panel and 6 members of a surgical subspecialty panel completed a 10-item questionnaire, discussed the measure on a moderated conference call, then completed the questionnaire again for their final ratings.

Validity testing results:

- Data element validity
  - The developer notes that most of the studies evaluating the data element validity of this measure examined data prior to 2010, when significant changes to the specifications were made to incorporate changes in the ICD-9-CM codes for DVT.
    - Nevertheless, these studies reported positive predictive values (PPVs) of 43% (95% CI, 34-53%) in VA hospitals, 44% (95% CI, 37-51%) in academic medical centers, and 48% (95% CI, 42-52%) in a national sample of volunteer hospitals. False-negative errors were extremely rare.
  - Based on AHRQ analyses and other studies in the peer-reviewed literature, an entirely new set of ICD-9-CM codes for superficial, upper extremity, and chronic venous thromboses were implemented. These codes are now excluded from the definition of PSI 12, which prompted AHRQ to reexamine the PPV of this indicator.
    - A number of studies have provided updated estimates:
      - A retrospective case-control study of risk factors for acute VTE after TKA in 15 teaching hospitals showed that the PPV of PSI 12 was 99% (125/126) and the negative predictive value (NPV) was 99.4% (460/463)
      - A chart abstraction data by 7 volunteer hospitals participating in AHRQ's Validation Pilot Project found an overall PPV of 81% (126/156), and most false positives were attributable to incomplete reporting of POA status (which now results in exclusion from the AHRQ reference population).
      - A study of FY 2012 data at a single academic medical center in Virginia found a PPV of 88% (95% CI: 80-93%).

- A study randomly sampling patients in 2007 and 2008 with PSI events from 3 Calgary hospitals found that the PPV for PSI12 was 90% (95% CI 67% to 99%).
- Several other findings are presented by the developer as well.

- Measure score validity:
  - The rates of PSI 12 events (v5.0) by day of index procedure are shown in a chart.
    - The developer states that index procedures occurring >2 days after admission may represent non-elective operations and suggest the possibility that the numerator event may have occurred before surgery (but after admission). Long preoperative delays may be a mutable process of care.
    - The developer observes that patients with long preoperative delays (4+ days) had 3 times higher risk of a postoperative VTE than patients who went to surgery in a more timely manner. The results suggest that reducing long preoperative delays may improve VTE rates. However, some of these preoperative delays may have been outside the provider's control, if these patients were too sick to undergo surgery before day 4 or did not require surgery until then.
  - For face validity, the developer reports that the multi-specialty Panel and Surgical Panel both rated the indicator as acceptable on overall usefulness as an indicator of potentially preventable complications of care.

Questions for the Committee:

- *Does the information provided by the developer demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The measure excludes the following cases:
  - with a principal ICD-9-CM or ICD-10-CM diagnosis code (or secondary diagnosis present on admission) for proximal deep vein thrombosis
  - with a principal ICD-9-CM or ICD-10-CM diagnosis code (or secondary diagnosis present on admission) for pulmonary embolism
  - where a procedure for interruption of vena cava occurs before or on the same day as the first operating room procedure*
  - any-listed ICD-9-CM or ICD-10-PCS procedure code for extracorporeal membrane oxygenation (ECMO)
  - any-listed ICD-9-CM or ICD-10-CM diagnosis code for acute brain or spinal injury present on admission
  - MDC 14 (pregnancy, childbirth, and puerperium)
  - with missing gender (SEX=missing), age (AGE=missing), quarter (DQTR=missing), year (YEAR=missing), or principal diagnosis (DX1=missing)
- To support these exclusions, the developer examined the percent of potential denominator cases excluded by each criterion as listed in the measure specifications.
  - The results of that analysis are provided in a table in the submission form.
- The developer notes that patients with a principal diagnosis of DVT or PE, or a secondary diagnosis of DVT or PE reported as present on admission, are excluded because the precipitating events happened before the targeted hospitalization, and are thus unlikely to reflect quality of care during that hospitalization.
- The other denominator exclusions are intended to reduce the number of flagged cases in which the diagnosis was non-preventable.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☐ **None** ☒ **Statistical model** ☐ **Stratification**

Conceptual rationale for SDS factors included ? ☒ **Yes** ☐ **No**

SDS factors included in risk model? ☐ **Yes** ☒ **No**

Risk adjustment summary

- The measure is expressed as a risk-adjusted rate per 1,000 surgical discharges; the risk-adjusted rate is computed using indirect standardization as the observed rate divided by the expected rate, multiplied by the reference population rate.
    - The observed rate is the number of discharge records where the patient experienced the PSI adverse event divided by the number of discharge records at risk for the event.
    - The expected rate is computed as the sum of the predicted value for each case divided by the number of cases for the unit of analysis of interest (i.e., hospital).
    - The predicted value for each case is computed using a hierarchical model (logistic regression with hospital random effect) and covariates for gender, age (in 5-year age groups, except for the youngest age range), Modified Diagnosis Related Groups, which are the base MS DRGs without any distinction for "comorbidity and complications" (CC/MCC), AHRQ Comorbidity Index, Major Diagnosis Categories (MDC) based on the principal diagnosis, and transfer in from another acute care hospital.
- Clinical risk-adjustment
    - For the risk-adjustment, the developer considered a standard set of covariates grouped into four categories: demographics, severity of illness, comorbidities and transfer-in status. Covariates that were considered as potential risk adjusters included gender and age, MDC, Modified Diagnostic Related Groups (MDRGs) (defined as the base MS-DRG without comorbidity or complication distinctions), AHRQ Comorbidity Software categories and whether they were transferred from another facility. Only those covariates with at least 30 cases for PSI 12 are retained. A parsimonious model was identified using backward stepwise selection with bootstrapping.
    - The measure's risk model includes 187 risk categories, including 26 age-gender categories in 5-year age categories between ages 30 and 89, and 2 age-gender categories ranging from below age 30 (i.e. 18-29) as one category and ages 90+ as another category, transfer in from another acute care facility and 13 comorbidities.
    - The remainder of selected risk factors account for the reason for admission and the type of surgery that was performed during the hospitalization, including MDC and MS-DRGs collapsed to remove Complication or Comorbidity/ Major Complication or Comorbidity (CC/MCC) distinctions.
    - The current risk adjustment coefficients for PSI 12 can be found in an Excel file attached to the submission.
    - To validate their risk-adjustment approach, the developer conducted an analysis to evaluate how strongly the risk adjustment model is associated with the event of interest.
        - The measure of discrimination, how well the risk adjustment model distinguishes events from non-events, is the **c-statistic** (also known as the area under a receiver operating characteristic curve).
            - The c-statistic is a measure of the proportion of discordant pairs of observations for which the observation with the event had a higher predicted probability from the risk-adjustment model than the observation without the event
        - The developer also evaluated the calibration of the risk adjustment model by evaluating how closely observed and predicted rates compare across deciles of the predicted rate. A model that is well calibrated will have observed values similar to predicted values across the predicted value deciles.
        - The results of this analysis are provided in a table in the submission form.
    - The developer's interpretation of their analysis is that the risk-adjustment model has moderately high

discrimination, based on a **c statistic of 0.751** (i.e., in 75% of randomly selected pairs of discordant observations, the patient who experienced PSI 12 had a higher probability of experiencing the event than the patient who did not).

- The developer also suggests that the measure is well calibrated, as the **observed to predicted ratio values across the deciles range between 0.87 to 1.09** for all deciles except the lowest decile

- SDS Adjustment
  - The developer notes that racial differences in the incidence of PSI12 – indeed most manifestations of venous thromboembolism – have been recognized for at least 15 years.
  - The developer suggests that these differences are thought by many to be the result of genetic predisposition, but notes that racial differences appear to be lower in the setting of nearly universal pharmacologic prophylaxis (e.g., total hip or knee arthroplasty, valve replacement) than in other clinical settings.
  - AHRQ is currently exploring the extent to which observed racial differences are consistent across states, and the extent to which adjusting for race would substantially change hospital-specific PSI 12 rates.
  - The developer states that there is no evidence or causal model to suggest that socioeconomic factors other than race are associated with postoperative thromboembolic events independent of quality of care, or are mediated by pre-hospital care (which may not fall within the proper realm of hospital accountability).
    - Accordingly, consistent with the guidance provided by NQF in the SDS Trial Period FAQs, AHRQ believes that it would be inappropriate to include other SDS variables in the risk-adjustment approach for PSI 12, which is an in-hospital outcome measure.

Questions for the Committee:

- *Is an appropriate risk-adjustment strategy included in the measure?*
- *Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented?*
- *Are all of the risk adjustment variables present at the start of care? If not, describe the rationale provided.*
- *Do you agree with the developer's rationale that there is no conceptual basis for adjusting this measure for SDS factors?*
- *Do you agree with the developer's decision, based on their analysis, to not include SDS factors in their risk-adjustment model?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- To determine if meaningful differences in performance measure scores among measured entities can be identified, the developer assesses the probability that a hospital is higher or lower than a benchmark or threshold, given hospital size.
- The developer suggests this analysis reflects whether the indicator can discriminate the best performing hospitals from the lower performing hospitals.
- The developer provides a table reporting the proportion of hospitals above (better than) and below (worse than) the Benchmark and Threshold rates and the proportion not classified as either above or below. The proportion of hospitals not classified as either better or worse have rates that fall within the 95% confidence interval.
- The developer's interpretation of their analysis is that, over all hospitals, this indicator has modest discrimination for identifying low or high performing hospitals; 35% of hospitals can be classified as better or worse than the threshold (the percentage classified as either above or below the threshold) and 33% better or worse than the benchmark (the percentage classified as either above or below the benchmark).
- The developer also notes that discrimination increases as hospital size increases.

Question for the Committee:

○ *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

N/A

2b7. Missing Data

- With regard to missing data, the developer reports that PSI 12 excludes cases with missing discharge disposition, age, sex, discharge quarter, discharge year, and principal diagnosis.
- The developer notes that for these variables, frequencies of missing data are typically less than 1% of the state database, suggesting it is unlikely the bias would occur from such a low frequency of missing data.
- The developer concludes that exclusion of cases with missing data for these variables is appropriate.

Guidance from the Validity Algorithm

[Box 1] Specifications consistent with evidence → [Box 2] Potential threats to validity addressed → [Box 3] Empirical validity testing conducted using the measure as specified → [Box 6] Testing conducted at the measure score level → [Box 7] Testing method described and appropriate → [Box 8] Moderate certainty or confidence that measure scores are valid → [Box 8b]

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

Committee pre-evaluation comments

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

2. Scientific Acceptability of Measure Properties

2a1. & 2b1. Specifications

Comments: ** Specifications are clear and consistent with the evidence.

2a2. Reliability Testing

Comments: **Measure score reliability testing was done using signal-to-noise analysis. The ratio was calculated at the hospital level and then summarized across the population of hospitals. Smaller hospitals have less reliable rates due to small denominators. Developer recommends using smoothed rates to account for the impact of hospitals with fewer than 100 qualifying discharges on the rate. Overall signal-to-noise ratio was 0.74 – moderate.

2b2. Validity Testing

Comments: **Both measure score and data element validity testing were done using empirical testing. Positive predictive values were found in studies of data element validity. False negatives were rare. Changes in coding prompted AHRQ to examine the PPV of the indicator -- positive predictive values were found.

The developer found that preoperative delays caused higher risk of postoperative VTE. For face validity, the multi-specialty and surgical panels both rated the indicator as acceptable on overall usefulness as an indicator of potentially preventable complications.

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: ** Exclusions are consistent with evidence and analysis.

A statistical model of risk adjustment is included. The interpretation of the analysis of the risk-adjustment model is that it has a moderately high discrimination based on a c-statistic of 0.751. The developer states that the measure is well-calibrated because the observed to predicted ratio values across the deciles range between 0.87 and 1.09. The developed indicates that the measure has modest discrimination for identifying low or high performing hospitals. Frequency of missing data are typically less than 1%

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.

- This measure is generated or collected by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)
- The required data elements are largely available in electronic health records or other electronic sources or existing electronic sources, a credible, near-term path to electronic collection is specified.
- ALL data elements are in defined fields in electronic claims.
- The indicator is based on readily available administrative billing and claims data.
- This version of the indicator requires present-on-admission (POA) data for risk-adjustment and for specification of the numerator and denominator.
- In 2007 POA indicators were added as data elements to the uniform bill form. A payment penalty was initiated on hospitals who did not include POA status on Medicare records beginning October 1, 2008.
- The developers' QI software has been publicly available at no cost since 2001; Users have over ten years of experience using the developers' QI software in SAS and Windows.
- There are no fees associated with this measure. Software is freely available from the developers Quality Indicators website.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*

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

Committee pre-evaluation comments

Criteria 3: Feasibility

3. Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: ** Data elements are electronic in claims database. Software is publicly available at not costs.

Criterion 4: Usability and Use

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

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

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

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

Accountability program details:

- Arizona Department of Health Services, AZ Hospital Compare, MONAHRQ website. Hospital quality ratings from all hospitals in Arizona.
- CareChex (Division of Quantros), Provides comprehensive reports of hospitals to consumers, providers and purchasers.
- Cigna - Centers of Excellence Hospital Value Tool – Health insurance company
- CMS Medicare Hospital Compare Program - Publically available database containing information about the quality of care at over 4,000 Medicare-certified hospitals across the U.S

- Colorado Hospital Association - Hospital quality ratings from hospitals in Colorado
- Commonwealth Fund, Why Not the Best - Provides performance and quality ratings for most US hospitals
- Connecticut Department of Health Services, CT Hospital Compare, MONAHRQ website - Hospital quality ratings from all hospitals in Connecticut.
- Connecticut Hospital Association - Provide quality of care for hospitals in Connecticut
- Florida Agency for Health Care Administration -Provide quality of care ratings from hospitals within Florida
- Healthcare Association of New York State - Supports availability of hospital quality and safety information to help patients make choices and assist providers in improving care
- HealthGrades - HealthGrades measures 40 million patient records from 4,500 hospitals nationwide for the most recent three-year period. Consumer-targeted hospital and provider ratings
- Hospital Safety Score - PSI 12 is one component of a single composite score that represents a hospital's overall performance in patient safety
- Illinois Department of Public Health - Provides access to information on hospital and safety data in hospitals in Illinois
- Iowa Healthcare Collaborative - Hospital quality ratings from hospitals in Iowa
- Kentucky Cabinet for Health and Family services - Hospital quality ratings from hospitals in Kentucky
- Kentucky Hospital Association Quality Data - Hospital quality ratings from most hospitals in Kentucky
- Louisiana Hospital Inform - Hospital quality ratings from hospitals in Louisiana
- Maine Health Data Organization (MHDO), MONAHRQ Website - Hospital quality ratings from all hospitals in Maine
- Maryland Health Care Commission, MONAHRQ Website - Collects and provides quality ratings on hospitals across Maryland
- Minnesota Community Measurement - Minnesota Community Measurement is a nonprofit healthcare data reporting organization. Provides quality ratings on hospitals across Minnesota.
- Nevada Compare Care, MONAHRQ website - Hospital quality ratings from most hospitals in Nevada
- Nevada Hospital Association - Transparency and Performance: Demonstrates Nevada hospitals activity relating to specific clinical indicators.
- New Jersey Department of Health - Public report of PSI performance for New Jersey Hospital
- Niagara Health Quality Coalition, New York State Hospital Report Card - Consumer focused public report of quality indicator performance for NY hospitals.
- Norton Healthcare - Report patient satisfaction scores in Norton Healthcare hospitals and their performance on nationally recognized quality indicators and practices.
- Oklahoma State Department of Health, MONAHRQ - Compares quality ratings on hospitals across Oklahoma
- South Dakota Association of Healthcare Organizations - Use PSI 12 in a composite of serious complications in report of Oregon hospital quality.
- Texas Health Resources - Provides quality and safety reports for all Texas Health Resources
- Think About It Colorado - Report hospital quality for all hospitals in Colorado
- U.S. News and World Report - National publication that lists ratings of U.S. medical centers based on performance
- Utah Department of Health, MONAHRQ website - Report hospital quality for all hospitals in Utah
- Virginia Health Information - Compares quality ratings on hospitals across Virginia
- Washington State, MONAHRQ website – Information system of inpatient care utilization, quality, and potentially avoidable stays in Washington State's community hospitals
- WHA Information Center (Wisconsin Hospital Association) - Wisconsin Inpatient Hospital Quality Indicators Report
- Quality Improvement (Internal to the specific organization) - Greenville Health System, Quality and Safety Report. All data was collected from four hospitals in the Greenville Health system and compared with internal rates
- Northwestern Memorial Hospital, Patient Safety Indicator Monitoring Plan - Quality improvement initiative at 894-bed academic hospital
- Upstate University Hospital - Report of hospital rates against national benchmark (published online)

- Quality Improvement (external benchmarking to multiple organizations) CMS Hospital Compare - Publically available performance measures for hospitals
- University HealthSystem Consortium/Vizient - Internal quality improvement efforts, documentation, and evaluation of AHRQ PSIs for quality improvement by its members

Improvement results

- During 2011-2013, nationwide rates of this measure have decreased from 4.0 to 3.7 cases/1,000 hospitalizations at risk.
- This decrease is consistent with the 43% decrease in hospital-associated VTE between 2010 and 2014 (i.e., an interval of 4 years versus 2 years) reported from the Medicare Patient Safety Monitoring System (MPSMS)

Unexpected findings (positive or negative) during implementation:

Potential harms:

- The developer discussed some potential unintended consequences related to this measure:
    - Although the developer reported no definite evidence of unintended consequences for this measure, several recent papers have focused on the problem of surveillance bias, or variation in the incidence of VTE across hospitals that may be attributable to screening and diagnostic practices.
    - These studies suggest that variation in testing practices may contribute to variation in PSI12 rates across hospitals, but it remains unclear whether these data reflect underdiagnoses of VTE at low-testing hospitals, over diagnosis at high-testing hospitals, or the true incidence of symptomatic VTE.
    - Use of PSI12 may inappropriately reward under-testing, but it may also appropriately penalize over-testing and over-diagnosis. Discouraging over-diagnosis and overtreatment of clinically insignificant VTE would be a desirable consequence of using PSI 12, because treatment of VTE is associated with a significant risk of hemorrhagic complications. Excluding distal DVTs, as AHRQ did in its V6 modification of the PSI12 numerator specification, may resolve this concern.

Feedback :

- Developer did not identify any specific feedback loops related to this 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

4. Usability and Use

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: ** "This measure is publicly reported and used in accountability programs in many states. Nationwide results (2011-13) decreased from 4.0 to 3.7 cases/1000 hospitalizations at risk, consistent with 43% decrease in hospital-associated VTE 2010-2014. Surveillance bias has been discussed as a possible problem that may be attirbutable to screening and diagnostic practices

Criterion 5: Related and Competing Measures

Related or competing measures

None

Harmonization
N/A

Pre-meeting public and member comments

Submitted by: Armstrong Institute for Patient Safety and Quality at Johns Hopkins University

We support efforts to measure patient safety in hospitals. We believe that valid and reliable measures of patient safety events are the foundation to improving performance and holding hospitals accountable.

Given the recent article by Winters et al. in Medical Care that found this measure did not meet validity thresholds when measured against the reference standard of a medical chart review, we would urge the standing committee to review the Medical Care article as part of their careful evaluation of the measure's validity.

Winters BD, Bharmal A, Wilson RF, Zhang A, Engineer L, Defoe D, Bass EB, Dy S, Pronovost PJ. Validity of the Agency for Health Care Research and Quality Patient Safety Indicators and the Centers for Medicare and Medicaid Hospital-acquired Conditions: A Systematic Review and Meta-Analysis. Medical care. 2016 Apr.

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0450
Measure Title: Perioperative Pulmonary Embolism or Deep Vein Thrombosis Rate (PSI 12)
IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: 0531 Patient Safety and Adverse Events Composite (PSI 90)
Date of Submission: 5/13/2016

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

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

Outcome

☒ Health outcome: Surgical Discharges with Pulmonary Embolism or Deep Vein Thrombosis

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

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

☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Deep vein thrombosis (DVT) is the formation of a blood clot in a deep vein—usually in the leg or pelvic veins. The most serious complication of a DVT is that the clot dislodges and travels to the lungs, becoming a pulmonary embolus (PE). Venous thromboembolism (VTE), which encompasses both DVT and PE, is common in the perioperative setting, especially after high-risk operations, and can be deadly. Clinical trials have demonstrated that mechanical and pharmacologic interventions can substantially reduce the risk of perioperative VTE among moderate and high-risk surgical patients, especially when these interventions are initiated before or immediately after surgery and continued until or after discharge. Case control studies have demonstrated that early ambulation after surgery can further reduce the risk of perioperative VTE among high-risk surgical patients who receive appropriate mechanical or pharmacologic prophylaxis. Effective and safe prophylactic measures are now available for most high risk patients, and numerous evidence-based guidelines have been published for the prevention of VTE (most notably by the American College of Chest Physicians and the American Academy of Orthopedic Surgeons).

As summarized in a 2015 AHRQ report on Preventing Hospital Associated Venous Thromboembolism (available at http://www.ahrq.gov/sites/default/files/wysiwyg/professionals/quality-patient-safety/patient-safety-resources/resources/vtguide/vteguide.pdf):
"Thromboprophylaxis for at-risk inpatients can reduce VTE by 30% to 65%, has a low incidence of major bleeding complications, and has well-documented cost-effectiveness… Numerous guidelines from authoritative bodies outlining appropriate use of thromboprophylaxis are available… yet study after study reflects unacceptably low rates of thromboprophylaxis in patients at risk… For example, a recent cross-sectional international study of almost 70,000 patients in 358 hospitals found that appropriate prophylaxis was administered in only 58.5% of surgical and 39.5% of medical inpatients at risk for VTE; another U.S. registry found only 42 percent of patients with hospital-associated DVT received prophylaxis within 30 days prior to diagnosis… This constellation of facts presents a powerful imperative for improvement."

"This "implementation gap" in VTE prophylaxis between evidence-based best practice and actual practice in the real world has not gone unnoticed as a major opportunity for improvement. In 2008, the U.S. Surgeon General produced a call-to-action document for VTE prevention… In addition, key goals for VTE prevention are in place from the National Quality Forum and the Joint Commission… VTE Prevention is one of the focus areas of the Partnership for Patients, a major effort from the Centers for Medicare & Medicaid Services (CMS) to foster accelerated improvement…. Reports commissioned by AHRQ called thromboprophylaxis the "number one" patient safety practice… and a 2013 update continues to list improved prophylaxis for VTE as a top 10 patient safety strategy to act on now…. The American Public Health Association has stated that the "disconnect between evidence and execution as it relates to DVT prevention amounts to a public health crisis…"

"Various strategies to improve the use of thromboprophylaxis have been demonstrated to be effective, including computerized order sets with electronic alerts, or pre-printed orders and quality improvement in the form of clinician education programs, audit, and feedback, but further efforts are required at improving the translation of data from clinical trials into clinical practice…" Use of the Patient Safety Indicator (PSI) Perioperative Pulmonary Embolism or Deep Vein Thrombosis Rate (PSI 12) and related measures developed by The Joint Commission, encourages providers to adopt the processes or structures of care of the best performing providers, and may empower consumers to select better performing providers or to adhere to recommended prophylactic modalities.

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☒ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.
Please note that this is an outcome measure, so a systematic review of the body of evidence that supports the performance measure is not required. However, information is provided in 1a.4.1, 1a.4.2, and 1a.8 below, to provide additional context and support for the measure.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

1. Guyatt GH, Eikelboom JW, Gould MK, et al. Approach to Outcome Measurement in the Prevention of Thrombosis in Surgical and Medical Patients: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest 2012;141(2_suppl):e185S-e194S. (Quoted below)

Additional Guidelines:

Lyman GH, Khorana AA, Kuderer NM, et al. American Society of Clinical Oncology Clinical Practice. Venous thromboembolism prophylaxis and treatment in patients with cancer: American Society of Clinical Oncology clinical practice guideline update. J Clin Oncol. 2013 Jun 10;31(17):2189-204)

American Academy of Orthopedic Surgeons:
http://www.aaos.org/Research/guidelines/VTE/VTE_full_guideline.pdf
http://www.aaos.org/Research/guidelines/HipFxSummaryofRecommendations.pdf

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

"Although some studies have limitations of lack of concealment and blinding, evidence from meta-analyses of randomized controlled trials (RCTs) strongly suggests that prophylaxis with an anticoagulant or aspirin reduces symptomatic VTE and fatal PE in medical and surgical patients. In patients undergoing orthopedic, general, or urological surgery, unfractionated heparin (UFH) reduces the risk of fatal PE by about two-thirds; in patients undergoing hip or knee arthroplasty or hip fracture surgery, vitamin K antagonists (VKAs) reduce the risk of symptomatic VTE by about four-fifths; in patients undergoing hip or knee arthroplasty, extended-duration low-molecular-weight heparin (LMWH) or warfarin reduces the risk of symptomatic VTE by about three-fifths; in medical patients at highest risk, UFH, LMWH, danaparoid, or fondaparinux reduces the risk of PE by about two- to three-fifths; and in patients undergoing abdominal or pelvic surgery, LMWH reduces the risk of symptomatic VTE by about four-fifths. Antiplatelet therapy also is effective for the prevention of VTE in the highest-risk surgical or medical patients, reducing the risk of PE by about one-half and DVT by about three-fifths. Similar relative risk reductions are seen in trials comparing the efficacy of anticoagulant prophylaxis with placebo or no treatment based on a surrogate outcome; compared with placebo or no treatment, prophylactic anticoagulants reduce the relative incidence of silent DVT diagnosed through screening venography by 30% to 70%... Collectively, the meta-analysis data indicate that prophylactic anticoagulants are effective for the prevention of patient-important VTE and that the benefit-risk trade-off justifies their use in patients who are at sufficiently high risk of symptomatic VTE." (pages e185S-e186S)

… "the compelling evidence of a decrease in fatal PE that exists for anticoagulants and for aspirin does not exist for mechanical methods (page e186S) ."[1]

Similar guidelines supporting the routine use of pharmacologic prophylaxis in selected populations have been published by several professional organizations:

1. For "all patients with malignant disease undergoing major surgical intervention," by the American Society of Clinical Oncology.
2. For "patients undergoing elective hip and knee arthroplasty," by the American Academy of Orthopedic Surgeons.
3. For "management of hip fractures in the elderly," by the American Academy of Orthopedic Surgeons.
4. For "high risk patients undergoing gynecologic surgery" (defined as "surgery lasting less than 30 minutes in patients older than 60 years or with additional risk factors; major surgery in patients older than 40 years or with additional risk factors"), by the American College of Obstetricians and Gynecologists.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

Not applicable

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

Not applicable

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

Not applicable

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

Not applicable

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

Not applicable

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

Not applicable

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

Not applicable

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

Not applicable

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Not applicable

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

Not applicable

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Not applicable

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

Not applicable

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

Not applicable

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

Not applicable

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

Not applicable

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: Click here to enter date range

Not applicable

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

Not applicable

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

Not applicable

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

Not applicable

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

Not applicable

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for each new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

Not applicable

1a.8 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.8.1 What process was used to identify the evidence?

Formal environmental scans of the literature, including routine PubMed searches are performed to continually update evidence. The current evidence review was conducted in January 2016. Search terms included relevant MeSH terms (Venous Thromboembolism or VTE, Pulmonary embolus (PE) or embolism, DVT or Thrombosis) with MeSH terms (patient admission, hospitals, inpatient, patient safety, AHRQ) to identify studies examining quality of inpatient care. The search was limited to English-language publications.

1a.8.2. Provide the citation and summary for each piece of evidence

Association with other adverse outcomes: Cost/LOS, Readmissions and Mortality

Although much of the evidence regarding cost/length of stay (LOS), readmission and mortality comes from studies that use earlier versions of Perioperative Pulmonary Embolism or Deep Vein Thrombosis Rate (PSI 12) which did not include present-on-admission information, it is believed that the overall results would not drastically change with updated versions of the indicator. Cases from the HCUP Nationwide Inpatient Sample that were flagged by this Patient Safety Indicator (PSI) in 2000 had 6.6% excess mortality, 5.4 days of excess hospitalization, and $21,709 in excess hospital charges, relative to carefully matched controls that were not flagged.[1] This finding was confirmed in the Veterans Affairs hospital system, where cases that were flagged by this PSI in 2001 had 6.1% excess mortality, 4.5-5.5 days of excess hospitalization, and $7,205-9,064 in excess hospital costs, relative to carefully matched controls that were not flagged.[2] Carey and Stefos re-estimated the financial impact of each PSI 12 event in the VA system in 2007 as $17,453-18,935, using more sophisticated cost accounting and econometric methods.[3] In another study based on HCUP SID from seven states in 2004 that permit linkage of serial hospitalizations, this indicator was associated with risk ratios of 1.35 for inpatient death, 1.28 for readmission within three months, and 1.25 for readmission within one month (after adjusting for age, gender, payer, comorbidities, specific surgical DRGs, and APR-DRG severity levels).[4] Similarly, in a multivariable analysis of Veterans Health Administration data from 2003-2007, hospitalizations with a PSI 12 event were 33% more likely to result in a readmission within 30 days (OR 1.33; 95% CI 1.23-1.44), after adjusting for age, sex, comorbidities, and other PSI events.[5]

Several other studies have focused on narrower clinical cohorts, with similar results. Bohensky et al. examined cost and length of stay (LOS) following complications in 139,031 knee arthroscopy cases in the Victorian Admitted Episodes Dataset (2000 to 2009). VTE events were the most common complication (0.3%) and the cumulative excess 30-day cost of VTE was $3227 (95% CI $3211-3244). Patients who experienced VTE also had longer median LOS (6 days vs. 1 day, $p<0.01$) than those without VTE.[6] Ramanan et al. used 2007-2009 National Surgical Quality Improvement Program (NSQIP) data on patients undergoing vascular surgery to show that VTE events increased overall mortality risk among patients with DVT (1.5% to 6.2%) or PE (1.5% to 5.7%), compared to those without VTE.[7] Using data from the NSQIP Semi Annual Reports for 197 US and Canadian hospitals (2007-2008), Borgi et al. demonstrated that VTE events were positively and statistically significantly associated with postoperative mortality (regression slope 0.393; 95% CI 0.235 to 0.551, $p<0.0001$).[8] In an analysis of Medicare claims data for patients undergoing any of 6 cancer resections in 2005-2009, Short et al. found that after adjusting for patient factors (age, sex, race, income), hospital factors (hospital volume, surgeon volume, surgeon specialty designation, hospital resources, patient characteristics) and tumor factors (tumor stage, site), costs increased significantly in association with postoperative VTE for all six types of surgery ($p<0.001$).[9] Based on an analysis of the 501,908 hospitalizations involving a brain tumor in the NIS between 2002 and 2010, Rahman et al. (2013) found that patients with postoperative DVT or PE had significantly longer length-of-stay, on average, than patients without these complications (10.4 vs 6.3 days and 8.8 vs. 6.4 days, respectively; $p < 0.0001$ for both).[10]

As noted in National Quality Forum (NQF) Measure Submission Form Usability and Use Section 4b. - Improvement, AHRQ recently reported a 43% national reduction in the incidence of postoperative VTE between 2010 and 2014, based on the Medicare Patient Safety Monitoring System (MPSMS), not PSI 12.[11] Using Umscheid et. al.'s method to calculate ranges for the annual number of preventable events, deaths and annual costs, 12,000 averted events led to 1,248 prevented deaths and a projected cost savings of $96 million in 2014. This proportion of potentially preventable PSI 12 events appears to exceed the proportion of potentially preventable deaths in 30-day mortality measures (estimated at 6%[12] to 27%[13]), and the proportion of potentially preventable readmissions in measures of 30-day readmissions (estimated at 23%[14] to 27%[15]), and is comparable to the 26% to 54% of surgical site infections considered potentially preventable.[16] Gidwani and Bhattacharya (2015) found that CMS payment reform was associated with a 35% lower incidence of hospital-acquired VTE among Medicare patients aged 65-69 years who had hip or knee arthroplasty.[17]

Association with Health System Characteristics

Several studies have examined the influence of various hospital and health system characteristics on the rate of postoperative PE and DVT. One study demonstrated that hospitals with higher percentages of registered nurses with baccalaureate or higher degrees had lower rates of PSI 12,[18] while studies were inconclusive regarding the impact of hospital factors such as being within the VA healthcare system,[2,19,20] teaching status,[2,21] bed size,[21] location,[21] nurse staffing hours,[19] safety climate[22] and the implementation of duty-hour regulations.[23] Another study found lower rates of postoperative PE and DVT at low procedure volume hospitals (compared to high-volume hospitals), rural hospitals (compared to urban hospitals), and non-teaching hospitals (compared to teaching hospitals), but statistical test values were not provided.[7]

Association with Processes of Care

Several recent studies examined the impact of efforts to improve VTE prophylaxis adherence, tracking changes in incidence over time as processes improved. Most of these studies reported favorable results, with the notable exception of cancer patients. For example, implementation of "mandated risk assessment" with computerized DVT prophylaxis order entry at a tertiary cancer center increased use of prophylaxis without reducing VTE incidence,[24] whereas similar protocols reduced the incidence of postoperative VTE on an vascular surgery service from 1.49% to 0.38%,[25] at a large Russian medical center from 0.88% to 0.42%,[26] and at a large medical center in Abu Dhabi from 0.9-3.1% to 0.1-0.2%.[27] Nelson et al. (2015) analyzed 2006-2011 surgical registry data on colorectal surgery from Washington and reported that use of in-hospital postoperative VTE chemoprophylaxis increased from 59.6% to 91.4%, but 90-day VTE rates did not decrease.[28] Heslin et al. reported that among 12 surgical services in a single institution the most common contributing factor for PSI 12 was "failure to follow protocol," but they did not report the impact of improved adherence on PSI 12.[29] Hussey et al tested an alpha version of the AHRQ QI Toolkit in a one-year quality improvement initiative at an academic medical center. After the electronic medical record was revised so that DVT prophylaxis would be a mandatory part of the order set, PSI12 rates decreased from 20.7 to 15.9.[30] A similar clinical decision support intervention at the University of Pennsylvania was associated with increased use of "recommended" prophylaxis (from 32.3% to 60.0%) and a concurrent drop in PSI 12 rates from 21.8 to 17.3.[31] The University of California recently reported that a five-campus collaborative effort to improve VTE risk stratification and prophylaxis achieved a 23.8% relative reduction in the incidence of PSI 12 in 2014 relative to 2011.[32] A similar program at Boston Medical Center, which also included an emphasis on early ambulation, was associated with an 84% decrease in DVT incidence (from 1.9% to 0.3%) and a 55% decrease in PE incidence (from 1.1% to 0.5%), lowering the observed-to-expected VTE ratio from 3.41 to 0.94.[33]

In a series of studies from John Hopkins, use of risk-appropriate VTE prophylaxis in surgical patients increased from 26% (42 of 161) at baseline to 68% (178 of 262) within 12 months, and to 85% after implementation of computer-based "smart order sets."[34] A retrospective review of 92 patients diagnosed with hospital-acquired VTE found that only 43 (47%) received defect-free care, while 49 (53%) had potentially preventable VTE. On the trauma service, 56.0% of residents prescribed "optimal, risk-appropriate" VTE prophylaxis, while attending physicians had a compliance rate of 74.2% (interquartile range, 72.6%-77.3%), indicating that resident practice variation may be an important contributor to VTE events at teaching hospitals.[35] Lau et al. (2015) reported that a performance feedback scorecard with individual peer-to-peer coaching increased the percentage of these residents providing defect-free care from 45% to 78% and reduced the incidence of postoperative VTE from 0.81% to 0.38-0.39%.[36]

AHRQ's Evidence-based Practice Review on Patient Safety summarized the state of the field: "Even though high quality evidence exists for safe and effective strategies to reduce the risk of VTE, studies continue to show that many hospitalized patients are not given risk-appropriate VTE prophylaxis. One recent study across 32 countries found that only 59% of at-risk surgical and 40% of at-risk medical patients received guideline-recommended VTE prophylaxis, and a United States registry study found that only 42% of patients diagnosed with DVT during a hospitalization had received prophylaxis…"[37] Similar findings have been reported from Europe[38] and from 28 Veterans Health Administration hospitals, where "accounting for contraindications and early VTE occurrence, a total of 78% of cases [with PSI 12] and 80% of controls [without PSI 12] were appropriately managed".[39]

Delayed Ambulation

Based on observational data from case control studies and longitudinal intervention studies, delayed ambulation is an independent risk factor for VTE after orthopedic surgery, even accounting for appropriate pharmacologic prophylaxis. In a case-control study of patients undergoing total knee arthroplasty (TKA) in 15 teaching hospitals, among PSI 12 cases with an objectively documented acute VTE within 9 days of surgery (N=130) and randomly selected controls (N=463), only 68% ambulated on day 1 or 2 after surgery despite all patients receiving thromboprophylaxis (pharmacologic in 80%, mechanical alone in 20%). Factors significantly associated with VTE (after adjusting for age, sex, history of VTE, and BMI) were bilateral TKA (OR=4.2; 95% CI: 1.9-9.1), receipt of pharmacological prophylaxis (OR=0.5; 95% CI: 0.3-0.8), and ambulation by postoperative day 2 (OR=0.3; 95% CI: 0.1-0.9).[40] In an earlier case control study based on a sampling frame with 25,388 Medicare fee-for-service beneficiaries who underwent unilateral total hip arthroplasty (THA) in any nonfederal hospital in California, White et al. compared processes of care between 297 randomly selected cases with VTE within 3 months after surgery and 592 randomly selected controls. Factors independently associated with VTE included initial ambulation before day 2 after surgery (OR=0.7; 95% CI 0.5–0.9), use of pneumatic compression (among patients with body-mass index <25; OR=0.3; 95% CI 0.2–0.6), and use of warfarin after discharge (OR=0.6; 95% CI 0.4–1.0).[41] These studies suggest a population fraction of post-arthroplasty VTE attributable to delayed ambulation of at least 10% and perhaps over 40%.

Two studies have reported single-center results of prospectively implementing early ambulation postoperative care protocols. Chandrasekaran et al. found that getting patients out of bed or walking for at least 15–30 minutes twice on the first day after TKA significantly reduced the odds of asymptomatic or symptomatic VTE (OR=0.35; 95% CI: 0.13-0.94) compared with the previous practice of confining patients to bed on that day.[42] Similarly, Pearse et al. implemented a treatment protocol that involved showering and walking up to 30 meters within 24 hours after TKA, and observed a substantial reduction in the odds of asymptomatic or symptomatic DVT (OR=0.04; 95% CI 0.004-0.30).[43] These findings are supported by several cohort studies summarized in a recent structured review.[44]

Association with Patient and Clinical Characteristics

Studies have shown variation in PE/DVT by procedure type, suggesting the importance of risk adjustment.[45, 46,47] Total operative time is also associated with increased VTE risk. Kim et al. (2015) reported that the risk of VTE in NSQIP data increased in a stepwise manner with the procedure standardized duration of general anesthesia time.[48] These findings were confirmed by Daley et al. (2015), using a measure of whether total operative time exceeded the upper 95% confidence limit of its expected value.[49]

Several studies have examined the association between patient characteristics and rates of pulmonary embolism and deep vein thrombosis. Associations between PSI12 and patient characteristics have been found for black race (for post-surgical DVT but not PE),[50] gender,[46,] age,[51,] obesity,[51] and select comorbidities (postoperative infection or stroke,[7] disseminated cancer,[7] dependent functional status,[7] return to operating room,[7] preoperative hyponatremia,[38] irritable bowel disease,[52] and congestive heart failure and cancer.[46] Other preoperative risk factors for VTE were identified in studies by Jamal et al. (2015)[53], Moghadamyeghaneh et al. (2014)[54], Martin et al. (2015)[55], Nelson et al. (2015)[56], Kimmell and Jahromi (2015)[57], Hoh et al. (2015)[58], Bekelis et al. (2015)[59], Swenson et al. (2015)[60], Wang et al. (2015)[61], and Greaves et al. (2015)[62] included: age, ASA risk classification (for colorectal surgery), white race (for esophageal surgery), body mass index (for hysterectomy and colorectal and bariatric surgery), cancer (for craniotomy and hysterectomy) and disseminated cancer (for colorectal surgery), chronic steroid use, emergent or non-elective surgery, open (versus laparoscopic) surgery (for colorectal and bariatric surgery), duration of pre-surgical hospitalization, preoperative sepsis, previous cardiac surgery, weight loss, hypoalbuminemia (for colorectal surgery), history of prior VTE, operation for inflammatory disease (for colorectal surgery), transfer from acute care hospital (for craniotomy), dependent functional status (for craniotomy), or individual comorbidities such as peripheral vascular disease and prior stroke (for craniotomy). Risk models have been developed and validated for VTE; Obi et al. (2015)[63] and Hachey et al. (2016)[64] validated the Caprini VTE risk assessment model among critically ill surgical patients and after lung cancer resection.

Some of the identified risk factors are at least partially under providers' control, and may account for some of the observed hospital-level variation in PSI12 rates (e.g., pre-surgical days, duration of general anesthesia or surgery, open versus laparoscopic approach, and postoperative complications such as prolonged mechanical ventilation and unplanned reintubation). Surgical duration is an especially noteworthy factor because of its association with resident involvement in surgery.

Validity of ICD coding

Several studies have studied the validity of administrative data in detecting postoperative PE/DVT. Most, including Winters et al.,[65] examine data prior to 2010 when significant changes to the indicator were made to incorporate changes in DVT codes that allowed the identification of clinically insignificant superficial and upper extremity clots.

Early studies of PSI 12 conducted in conjunction with the AHRQ Validation Pilot Project estimated the positive predictive value (PPV) of PSI12 (v3.1) to be 48% (95% CI 42%-67% in a national sample of volunteer hospitals.[66] Among 112 randomly selected cases in 28 acute care VA hospitals, the PPV for PSI 12 was 43% (95% CI 34 to 53) compared to nurse abstracted records.[67] At least four other studies assessed PSI 12 relative to clinical registries that track postoperative venous thrombosis, such as the National Surgical Quality

Improvement Program (NSQIP) and the Veterans Health Administration Surgical Quality Improvement Program (VASQIP). However, the definition of postoperative venous thrombosis in NSQIP and VASQIP is based on a positive (or "high probability") imaging test with "initiation of anticoagulation therapy" and/or "placement of mechanical interruption" in the inferior vena cava, whereas the AHRQ definition is based on a physician-documented diagnosis of PE/DVT (without regard to whether the condition was treated). After linking 55,752 hospitalizations in the 2001 VA inpatient administrative data with the corresponding VASQIP records, Romano et al. reported that the version 3.0 specification of PSI 12 had a sensitivity of 56% (95% CI, 50-63%) and a PPV of 22% (95% CI, 19-25%).[68] Mull et al. replicated this study with version 4.1 PSI software and 268,771 records from 2003-2007, and reported that PSI 12 had a sensitivity of 65% (95% CI, 63-67%) and PPV of 31% (95% CI, 30-33%).[69] However, when these authors reviewed 20 discrepancies manually (i.e., PSI 12 positive, VASQIP negative), 14 were found to be true positives by PSI 11 and false negatives in VASQIP. This finding suggests that the PPV of PSI 12 is substantially higher than 31%; in other words, VASQIP cannot be regarded as a gold standard. Single center studies by Cima et al. and Koch et al. also explored disagreement between PSI 12 and NSQIP, but the findings were not reported in sufficient detail to explain observed discrepancies (e.g., kappa=0.60, 95% CI 0.52-0.67; sensitivity=58%; PPV=42%).[70,71] Another study focused on patients undergoing pancreaticdudenectomy, finding discordance (0.4% NIS vs. 2.2% NSQIP, p<0.001)[31,72]

These studies are now of limited historical interest, because the advent of POA coding, the implementation of more specific ICD-9-CM codes for VTE, and AHRQ's modification of the PSI 12 numerator specification to take advantage of these coding changes have markedly improved the validity of PSI 12. Sadeghi et al. used two sources of data from 2009-2010 to generate updated estimates: (1) the UHC retrospective case-control study of risk factors for acute VTE after TKA in 15 teaching hospitals; (2) a chart abstraction data by 7 volunteer hospitals participating in AHRQ's Validation Pilot Project.[73] In the UHC sample, the PPV of PSI 12 was 99% (125/126) and the negative predictive value (NPV) was 99.4% (460/463). In the AHRQ sample, the overall PPV was 81% (126/156) and most false positives were attributable to incomplete reporting of POA status. Ramanathan et al. similarly reported a PPV of 88% (95% CI: 80-93%) from FY 2012 data at a single academic medical center in Virginia.[74] From Johns Hopkins Hospital in 2010-2011, Lau et al. reported a PPV of 93% based on coding criteria and 83% based on clinical criteria (i.e., excluding catheter-associated thrombi).[75] Finally, Quan et al. randomly sampled patients in 2007 and 2008 with PSI events from 3 Calgary hospitals; the PPV for PSI12 was 90% (95% CI 67% to 99%).[76] In summary, these four studies from after 2010 indicate that the previously documented problem with the PPV of PSI 12 has largely resolved.

Surveillance bias

Surveillance bias remains a concern in the surgical community, with evidence from both administrative and clinical registry data sets. Bilimoria et al. investigated surveillance bias using Hospital Compare data, American Hospital Association (2010) survey data, and 2009-2010 Medicare claims data, finding that greater hospital adherence to VTE prophylaxis had a weakly negative association with risk-adjusted VTE event rates (r^2=4.2%, p=0.03) and risk-adjusted VTE rates increased concordantly with VTE imaging use rates (p<0.001).[77] Holcomb et al. (2015) studied 25,975 patients meeting the criteria for the Surgical Care Improvement Project (SCIP)-VTE measures at 79 VA facilities and reported a positive correlation between inpatient surveillance and inpatient VTE rates at the hospital level (R=0.33, P=.003) but no significant correlation of inpatient surveillance with either post-discharge surveillance (R=0.11, P=.29) or post-discharge VTE rates (R=0.03, P=.76).[78] Ju et al. used NSQIP data to identify VTE events and Medicare claims data to obtain information about use of VTE imaging; mean risk-adjusted VTE rates (within 30 days after surgery) were significantly lower in hospitals in the lowest quartile of VTE imaging use (1.13%) than in hospitals in the highest quartile (1.92%, p<0.001).[79] Similarly, Pierce et al. showed in the National Trauma Data Bank, with 147 hospitals from 2001-2005, that "hospitals with an ultrasound rate ≤2% had a 1.07% (95% CI: 1.05-1.09%) increase in reported DVT rate for every 1% increase in ultrasound rate."[80,81] Studies have not examined whether the observed data reflect underdiagnosis of VTE at low-testing hospitals, overdiagnosis of VTE at high-testing hospitals, or the underlying true incidence of symptomatic VTE, although diagnostic practices may represent the most plausible explanation.[82,83]

Overdiagnosis of VTE among asymptomatic or minimally symptomatic patients may lead to overtreatment, with the known adverse effects of anticoagulation and/or inferior vena cava (IVC) device placement. Evidence-based guidelines note that "although distal DVT may be present in patients with a normal proximal ultrasound, it is seldom if ever associated with important clinical sequelae."[84] At least one large randomized controlled trial showed that sonography limited to proximal veins is just as safe as whole-leg ultrasound, because distal thromboses generally do not require treatment.[85] To minimize the impact of overdiagnosis of clinically unimportant distal thromboses on hospital-specific PSI 12 rates, AHRQ removed isolated thrombosis of calf veins (ICD-9-CM 453.42) from the V6 specification reviewed by the NQF Patient Safety Steering Committee in 2015. However, pulmonary embolism may also be overdiagnosed by reading small subsegmental filling defects as pulmonary emboli (rather than as "small sub-segmental filling defects of undetermined significance", which is a more appropriate term).[86] This problem of overdiagnosis and overtreatment of clinically unimportant VTE has received increasing attention in the clinical literature.[87,88] Given the negative economic and health consequences of being labeled as having VTE, reducing overdiagnosis may improve the overall health of the population.

References

1. Zhan C, Miller MR. Administrative data based patient safety research: a critical review. *Qual Saf Health Care.* 2003;12 Suppl 2:ii58-63.
2. Rivard PE, Christiansen CL, Zhao S, Elixhauser A, Rosen AK. Advances in Patient Safety: Is There an Association Between Patient Safety Indicators and Hospital Teaching Status? In: Henriksen K, Battles JB, Keyes MA, Grady ML, eds. *Advances in Patient Safety: New Directions and Alternative Approaches (Vol. 2: Culture and Redesign)*. Rockville (MD)2008.
3. Carey K, Stefos T. Measuring the cost of hospital adverse patient safety events. *Health economics.* 2011;20(12):1417-1430.
4. Friedman B, Encinosa W, Jiang HJ, Mutter R. Do patient safety events increase readmissions? *Med Care.* 2009;47(5):583-590.
5. Rosen AK, Loveland S, Shin M, et al. Examining the impact of the AHRQ Patient Safety Indicators (PSIs) on the Veterans Health Administration: the case of readmissions. *Med Care.* 2013;51(1):37-44.
6. Bohensky MA, Ademi Z, deSteiger R, et al. Quantifying the excess cost and resource utilisation for patients with complications associated with elective knee arthroscopy: a retrospective cohort study. *Knee.* 2014;21(2):491-496.
7. Ramanan B, Gupta PK, Sundaram A, et al. In-hospital and postdischarge venous thromboembolism after vascular surgery. *J Vasc Surg.* 2013;57(6):1589-1596.
8. Borgi J, Rubinfeld I, Ritz J, Jordan J, Velanovich V. The differential effects of intermediate complications with postoperative mortality. *Am Surg.* 2013;79(3):261-266.
9. Short MN, Aloia TA, Ho V. The influence of complications on the costs of complex cancer surgery. *Cancer.* 2014;120(7):1035-1041.
10. Rahman M, Neal D, Fargen KM, Hoh BL. Establishing standard performance measures for adult brain tumor patients: a Nationwide Inpatient Sample database study. *Neuro Oncol.* 2013;15(11):1580-1588.
11. AHRQ. Saving Lives and Saving Money: Hospital-Acquired Conditions Update Interim Data from National Efforts To Make Care Safer 2010-2014. 2014.
12. Hayward RA, Hofer TP. Estimating hospital deaths due to medical errors: preventability is in the eye of the reviewer. *Jama.* 2001;286(4):415-420.
13. Dubois RW, Brook RH. Preventable deaths: who, how often, and why? *Ann Intern Med.* 1988;109(7):582-589.
14. van Walraven C, Jennings A, Forster AJ. A meta-analysis of hospital 30-day avoidable readmission rates. *J Eval Clin Pract.* 2012;18(6):1211-1218.
15. van Walraven C, Bennett C, Jennings A, Austin PC, Forster AJ. Proportion of hospital readmissions deemed avoidable: a systematic review. *Cmaj.* 2011;183(7):E391-402.
16. Umscheid CA, Mitchell MD, Doshi JA, Agarwal R, Williams K, Brennan PJ. Estimating the proportion of healthcare-associated infections that are reasonably preventable and the related mortality and costs. *Infect Control Hosp Epidemiol.* 2011;32(2):101-114.
17. Gidwani R, Bhattacharya J. CMS reimbursement reform and the incidence of hospital-acquired pulmonary embolism or deep vein thrombosis. *J Gen Intern Med.* 2015;30(5):588-596.
18. Blegen MA, Goode CJ, Park SH, Vaughn T, Spetz J. Baccalaureate education in nursing and patient outcomes. *J Nurs Adm.* 2013;43(2):89-94.
19. Rivard PE, Elixhauser A, Christiansen CL, Shibei Z, Rosen AK. Testing the association between patient safety indicators and hospital structural characteristics in VA and nonfederal hospitals. *Med Care Res Rev.* 2010;67(3):321-341.
20. Chen Q, Hanchate A, Shwartz M, et al. Comparison of the Agency for Healthcare Research and Quality Patient Safety Indicator Rates Among Veteran Dual Users. *Am J Med Qual.* 2013;29(4):335-343.
21. Rosen AK, Zhao S, Rivard P, et al. Tracking rates of Patient Safety Indicators over time: lessons from the Veterans Administration. *Med Care.* 2006;44(9):850-861.
22. Rosen AK, Singer S, Shibei Z, Shokeen P, Meterko M, Gaba D. Hospital safety climate and safety outcomes: is there a relationship in the VA? *Med Care Res Rev.* 2010;67(5):590-608.
23. Shelton J, Kummerow K, Phillips S, et al. Patient safety in the era of the 80-hour workweek. *J Surg Educ.* 2014;71(4):551-559.
24. Kukar M, Asaro J, Aquino A, Groman A, Skitzki J, Kane JM. Incidence of Venous Thromboembolic Events in Mandated Risk Assessment versus Optional DVT Prophylaxis Era at a Large Tertiary Cancer Center. *Am Surg.* 2015;81(9):893-898.
25. Durinka JB, Hecht TE, Layne AJ, et al. Aggressive venous thromboembolism prophylaxis reduces VTE events in vascular surgery patients. *Vascular.* 2015.
26. Nazarenko GI, Kleymenova EB, Payushik SA, Otdelenov VA, Sychev DA, Yashina LP. Decision support systems in clinical practice: The case of venous thromboembolism prevention. *Int J Risk Saf Med.* 2015;27 Suppl 1:S104-105.
27. Nimeri AA, Gamaleldin MM, McKenna KL, Turrin NP, Mustafa BO. Reduction of Venous Thromboembolism in Surgical Patients Using a Mandatory Risk-Scoring System: 5-Year Follow-Up of an American College of Surgeons National Surgical Quality Improvement Program. *Clin Appl Thromb Hemost.* 2015.
28. Nelson DW, Simianu VV, Bastawrous AL, et al. Thromboembolic Complications and Prophylaxis Patterns in Colorectal Surgery. *JAMA Surg.* 2015;150(8):712-720.
29. Heslin MJ, Taylor B, Hawn MT, et al. A 100% departmental mortality review improves observed-to-expected mortality ratios and University HealthSystem Consortium rankings. *J Am Coll Surg.* 2014;218(4):554-562.

30. Hussey PS, Burns RM, Weinick RM, Mayer L, Cerese J, Farley DO. Using a hospital quality improvement toolkit to improve performance on the AHRQ quality indicators. *Jt Comm J Qual Patient Saf.* 2013;39(4):177-184.
31. Umscheid CA, Hanish A, Chittams J, Weiner MG, Hecht TE. Effectiveness of a novel and scalable clinical decision support intervention to improve venous thromboembolism prophylaxis: a quasi-experimental study. *BMC medical informatics and decision making.* 2012;12:92.
32. (IEC) C-IEC, Ong M, Romano P, et al. Report: UC Center for Health Quality and Innovation Review and Assessment of Grant Impact to Date. 2014.
33. Cassidy MR, Rosenkranz P, McAneny D. Reducing postoperative venous thromboembolism complications with a standardized risk-stratified prophylaxis protocol and mobilization program. *J Am Coll Surg.* 2014;218(6):1095-1104.
34. Jiang J, Teng Y, Fan Z, Khan S, Xia Y. Does obesity affect the surgical outcome and complication rates of spinal surgery? A meta-analysis. *Clin Orthop Relat Res.* 2014;472(3):968-975.
35. Nguyen GC, Bernstein CN, Bitton A, et al. Consensus statements on the risk, prevention, and treatment of venous thromboembolism in inflammatory bowel disease: Canadian Association of Gastroenterology. *Gastroenterology.* 2014;146(3):835-848.e836.
36. Lau BD, Arnaoutakis GJ, Streiff MB, et al. Individualized Performance Feedback to Surgical Residents Improves Appropriate Venous Thromboembolism Prophylaxis Prescription and Reduces Potentially Preventable VTE: A Prospective Cohort Study. *Ann Surg.* 2015.
37. Shekelle PG, Wachter RM, Pronovost PJ, et al. Making health care safer II: an updated critical analysis of the evidence for patient safety practices. *Evid Rep Technol Assess (Full Rep).* 2013(211):1-945.
38. Markovic-Denic L, Zivkovic K, Lesic A, Bumbasirevic V, Dubljanin-Raspopovic E, Bumbasirevic M. Risk factors and distribution of symptomatic venous thromboembolism in total hip and knee replacements: prospective study. *Int Orthop.* 2012;36(6):1299-1305.
39. Borzecki AM, Cowan AJ, Cevasco M, et al. Is development of postoperative venous thromboembolism related to thromboprophylaxis use? A case-control study in the Veterans Health Administration. *Jt Comm J Qual Patient Saf.* 2012;38(8):348-358.
40. Sadeghi B, Romano PS, Maynard G, et al. Mechanical and suboptimal pharmacologic prophylaxis and delayed mobilization but not morbid obesity are associated with venous thromboembolism after total knee arthroplasty: a case-control study. *J Hosp Med.* 2012;7(9):665-671.
41. White RH, Gettner S, Newman JM, Trauner KB, Romano PS. Predictors of rehospitalization for symptomatic venous thromboembolism after total hip arthroplasty. *N Engl J Med.* 2000;343(24):1758-1764.
42. Chandrasekaran S, Ariaretnam SK, Tsung J, Dickison D. Early mobilization after total knee replacement reduces the incidence of deep venous thrombosis. *ANZ J Surg.* 2009;79(7-8):526-529.
43. Pearse EO, Caldwell BF, Lockwood RJ, Hollard J. Early mobilisation after conventional knee replacement may reduce the risk of postoperative venous thromboembolism. *J Bone Joint Surg Br.* 2007;89(3):316-322.
44. Zhang ZH, Shen B, Yang J, Zhou ZK, Kang PD, Pei FX. Risk factors for venous thromboembolism of total hip arthroplasty and total knee arthroplasty: a systematic review of evidences in ten years. *BMC Musculoskelet Disord.* 2015;16:24.
45. White RH, Zhou H, Romano PS. Incidence of symptomatic venous thromboembolism after different elective or urgent surgical procedures. *Thrombosis and haemostasis.* 2003;90(3):446-455.
46. Assareh H, Chen J, Ou L, Hollis SJ, Hillman K, Flabouris A. Rate of venous thromboembolism among surgical patients in Australian hospitals: a multicentre retrospective cohort study. *BMJ Open.* 2014;4(10):e005502.
47. Aziz F, Patel M, Ortenzi G, Reed AB. Incidence of Postoperative Deep Venous Thrombosis Is Higher among Cardiac and Vascular Surgery Patients as Compared with General Surgery Patients. *Annals of vascular surgery.* 2015;29(4):661-669.
48. Kim HK, Jung H, Cho J, Huh S, Lee JM, Kim YW. Therapeutic outcomes and thromboembolic events after treatment of acute arterial thromboembolism of the upper extremity. *Ann Vasc Surg.* 2015;29(2):303-310.
49. Daley BJ, Cecil W, Clarke PC, Cofer JB, Guillamondegui OD. How slow is too slow? Correlation of operative time to complications: an analysis from the Tennessee Surgical Quality Collaborative. *J Am Coll Surg.* 2015;220(4):550-558.
50. Adelani MA, Archer KR, Song Y, Holt GE. Immediate complications following hip and knee arthroplasty: does race matter? *J Arthroplasty.* 2013;28(5):732-735.
51. Streiff MB, Brady J, Grant AM, Grosse SD, Wong B, Popovic T. CDC Grand Rounds: Preventing Hospital-Associated Venous Thromboembolism *Morbidity and Mortality Weekly Report* 2014;63(90):190-193.
52. Nguyen DL, Parekh N, Bechtold ML, Jamal MM. National Trends and In-Hospital Outcomes of Adult Patients With Inflammatory Bowel Disease Receiving Parenteral Nutrition Support. *JPEN J Parenter Enteral Nutr.* 2014.
53. Jamal MH, Corcelles R, Shimizu H, et al. Thromboembolic events in bariatric surgery: a large multi-institutional referral center experience. *Surg Endosc.* 2015;29(2):376-380.
54. Moghadamyeghaneh Z, Hanna MH, Carmichael JC, Nguyen NT, Stamos MJ. A nationwide analysis of postoperative deep vein thrombosis and pulmonary embolism in colon and rectal surgery. *J Gastrointest Surg.* 2014;18(12):2169-2177.

55. Martin JT, Mahan AL, Ferraris VA, et al. Identifying Esophagectomy Patients at Risk for Predischarge Versus Postdischarge Venous Thromboembolism. *Ann Thorac Surg.* 2015;100(3):932-938; discussion 938.
56. Nelson RE, Grosse SD, Waitzman NJ, et al. Using multiple sources of data for surveillance of postoperative venous thromboembolism among surgical patients treated in Department of Veterans Affairs hospitals, 2005-2010. *Thromb Res.* 2015;135(4):636-642.
57. Kimmell KT, Jahromi BS. Clinical factors associated with venous thromboembolism risk in patients undergoing craniotomy. *J Neurosurg.* 2015;122(5):1004-1011.
58. Hoh DJ, Rahman M, Fargen KM, Neal D, Hoh BL. Establishing standard hospital performance measures for cervical spinal trauma: a Nationwide In-patient Sample study. *Spinal Cord.* 2015.
59. Bekelis K, Kalakoti P, Nanda A, Missios S. A Predictive Model of Unfavorable Outcomes After Benign Intracranial Tumor Resection. *World Neurosurg.* 2015;84(1):82-89.
60. Swenson CW, Berger MB, Kamdar NS, Campbell DA, Jr., Morgan DM. Risk factors for venous thromboembolism after hysterectomy. *Obstet Gynecol.* 2015;125(5):1139-1144.
61. Wang L, Pryor AD, Altieri MS, et al. Perioperative rates of deep vein thrombosis and pulmonary embolism in normal weight vs obese and morbidly obese surgical patients in the era post venous thromboembolism prophylaxis guidelines. *Am J Surg.* 2015;210(5):859-863.
62. Greaves SW, Holubar SD. Preoperative Hospitalization Is Independently Associated With Increased Risk for Venous Thromboembolism in Patients Undergoing Colorectal Surgery: A National Surgical Quality Improvement Program Database Study. *Dis Colon Rectum.* 2015;58(8):782-791.
63. Obi AT, Pannucci CJ, Nackashi A, et al. Validation of the Caprini Venous Thromboembolism Risk Assessment Model in Critically Ill Surgical Patients. *JAMA Surg.* 2015;150(10):941-948.
64. Hachey KJ, Hewes PD, Porter LP, et al. Caprini venous thromboembolism risk assessment permits selection for postdischarge prophylactic anticoagulation in patients with resectable lung cancer. *J Thorac Cardiovasc Surg.* 2016;151(1):37-44.e31.
65. Winters BD, Bharmal A, Wilson RF, et al. Validity of the Agency for Health Care Research and Quality Patient Safety Indicators and the Centers for Medicare and Medicaid Hospital-acquired Conditions: A Systematic Review and Meta-Analysis. *Medical Care.* 2016.
66. White RH, Sadeghi B, Tancredi DJ, et al. How valid is the ICD-9-CM based AHRQ patient safety indicator for postoperative venous thromboembolism? *Medical care.* 2009;47(12):1237-1243.
67. Kaafarani HM, A.M. B, K.M. I, et al. Validity of selected Patient Safety Indicators: opportunities and concerns. *J Am Coll Surg. .* 2011;212(6):924-934.
68. Romano PS, Mull HJ, Rivard PE, et al. Validity of selected AHRQ patient safety indicators based on VA National Surgical Quality Improvement Program data. *Health services research.* 2009;44(1):182-204.
69. Mull HJ, Borzecki AM, Loveland S, et al. Detecting adverse events in surgery: comparing events detected by the Veterans Health Administration Surgical Quality Improvement Program and the Patient Safety Indicators. *Am J Surg.* 2014;207(4):584-595.
70. Cima RR, Lackore K, Nehring S, et al. How best to measure surgical quality? Comparison of the Agency for Healthcare Research and Quality Patient Safety Indicators (AHRQ-PSI) and the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) postoperative adverse events at a single institution. *Surgery.* 2011;150(5):943-949.
71. Koch CG, Li L, Hixson E, Tang A, Phillips S, Henderson JM. What are the real rates of postoperative complications: elucidating inconsistencies between administrative and clinical data sources. *J Am Coll Surg.* 2012;214(5):798-805.
72. Enomoto LM, Hollenbeak CS, Bhayani NH, Dillon PW, Gusani NJ. Measuring surgical quality: a national clinical registry versus administrative claims data. *J Gastrointest Surg.* 2014;18(8):1416-1422.
73. Sadeghi B, White RH, Maynard G, et al. Improved Coding of Postoperative Deep Vein Thrombosis and Pulmonary Embolism in Administrative Data (AHRQ Patient Safety Indicator 12) After Introduction of New ICD-9-CM Diagnosis Codes. *Med Care.* 2013.
74. Ramanathan R, Leavell P, Wolfe LG, Duane TM. Agency for Healthcare Research and Quality patient safety indicators and mortality in surgical patients. *Am Surg.* 2014;80(8):801-804.
75. Lau BD, Haut ER, Hobson DB, et al. ICD-9 Code-Based Venous Thromboembolism Performance Targets Fail to Measure Up. *American journal of medical quality : the official journal of the American College of Medical Quality.* 2015.
76. Quan H, Eastwood C, Cunningham CT, et al. Validity of AHRQ patient safety indicators derived from ICD-10 hospital discharge abstract data (chart review study). *BMJ Open.* 2013;3(10):e003716.
77. Bilimoria KY, Chung J, Ju MH, et al. Evaluation of surveillance bias and the validity of the venous thromboembolism quality measure. *Jama.* 2013;310(14):1482-1489.
78. Holcomb CN, DeRussy A, Richman JS, Hawn MT. Association Between Inpatient Surveillance and Venous Thromboembolism Rates After Hospital Discharge. *JAMA Surg.* 2015;150(6):520-527.
79. Ju MH, Chung JW, Kinnier CV, et al. Association between hospital imaging use and venous thromboembolism events rates based on clinical data. *Ann Surg.* 2014;260(3):558-564; discussion 564-556.

80. Pierce CA, Haut ER, Kardooni S, et al. Surveillance bias and deep vein thrombosis in the national trauma data bank: the more we look, the more we find. *The Journal of trauma.* 2008;64(4):932-936; discussion 936-937.
81. Haut ER, Chang DC, Pierce CA, et al. Predictors of posttraumatic deep vein thrombosis (DVT): hospital practice versus patient factors-an analysis of the National Trauma Data Bank (NTDB). *The Journal of trauma.* 2009;66(4):994-999; discussion 999-1001.
82. Chung JW, Ju MH, Kinnier CV, Sohn MW, Bilimoria KY. Postoperative venous thromboembolism outcomes measure: analytic exploration of potential misclassification of hospital quality due to surveillance bias. *Ann Surg.* 2015;261(3):443-444.
83. Minami CA, Bilimoria KY. Are Higher Hospital Venous Thromboembolism Rates an Indicator of Better Quality?: Evaluation of the Validity of a Hospital Quality Measure. *Adv Surg.* 2015;49:185-204.
84. Bates S, Jaeschke, R; Stevens, S., Goodacre, S., Wells, P., Stevenson, M., Kearon, C., Schunemann, H., Crowther, M., Pauker, S., Makdissi, R., Guyatt, G. Diagnosis of DVT: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *ANTITHROMBOTIC THERAPY AND PREVENTION OF THROMBOSIS.* 2012;9TH ED: ACCP GUIDELINES:e351S-e418S.
85. Bernardi E, Camporese G, Buller HR, et al. Serial 2-point ultrasonography plus D-dimer vs whole-leg color-coded Doppler ultrasonography for diagnosing suspected symptomatic deep vein thrombosis: a randomized controlled trial. *Jama.* 2008;300(14):1653-1659.
86. Suh JM, Cronan JJ, Healey TT. Dots are not clots: the over-diagnosis and over-treatment of PE. *Emergency radiology.* 2010;17(5):347-352.
87. Righini M, Le Gal G, Bounameaux H. Venous thromboembolism diagnosis: unresolved issues. *Thrombosis and haemostasis.* 2015;113(6):1184-1192.
88. Wiener RS, Schwartz LM, Woloshin S. Time trends in pulmonary embolism in the United States: evidence of overdiagnosis. *Archives of internal medicine.* 2011;171(9):831-837.

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
PSI12_Measure_Evidence_Form_160513_v2.docx

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., the benefits or improvements in quality envisioned by use of this measure)*
Deep vein thrombosis (DVT) is the formation of a blood clot in a deep vein—usually in the leg or pelvic veins. The most serious complication of a proximal DVT is that the clot dislodges and can travel to the lungs, becoming a pulmonary embolus (PE). Venous thromboembolism (VTE) is common in the perioperative setting, especially after high-risk operations, and can be deadly. Clinical trials have demonstrated that mechanical and pharmacologic interventions can substantially reduce the risk of perioperative VTE among moderate and high-risk surgical patients, especially when these interventions are initiated before or immediately after surgery and continued until or after discharge. Case control studies have demonstrated that early ambulation after surgery can further reduce the risk of perioperative VTE among high-risk surgical patients who receive appropriate mechanical or pharmacologic prophylaxis. Effective and safe prophylactic measures are now available for most high risk patients, and numerous evidence-based guidelines have been published for the prevention of VTE (most notably by the American College of Chest Physicians and the American Academy of Orthopedic Surgeons).

As summarized in a 2015 AHRQ report on Preventing Hospital Associated Venous Thromboembolism (available at http://www.ahrq.gov/sites/default/files/wysiwyg/professionals/quality-patient-safety/patient-safety-resources/resources/vtguide/vteguide.pdf):
"Thromboprophylaxis for at-risk inpatients can reduce VTE by 30% to 65%, has a low incidence of major bleeding complications, and has well-documented cost-effectiveness… Numerous guidelines from authoritative bodies outlining appropriate use of thromboprophylaxis are available… yet study after study reflects unacceptably low rates of thromboprophylaxis in patients at risk…

For example, a recent cross-sectional international study of almost 70,000 patients in 358 hospitals found that appropriate prophylaxis was administered in only 58.5% of surgical and 39.5% of medical inpatients at risk for VTE; another U.S. registry found only 42 percent of patients with hospital-associated DVT received prophylaxis within 30 days prior to diagnosis… This constellation of facts presents a powerful imperative for improvement."

"This "implementation gap" in VTE prophylaxis between evidence-based best practice and actual practice in the real world has not gone unnoticed as a major opportunity for improvement. In 2008, the U.S. Surgeon General produced a call-to-action document for VTE prevention… In addition, key goals for VTE prevention are in place from the National Quality Forum and the Joint Commission… VTE Prevention is one of the focus areas of the Partnership for Patients, a major effort from the Centers for Medicare & Medicaid Services (CMS) to foster accelerated improvement…. Reports commissioned by AHRQ called thromboprophylaxis the "number one" patient safety practice… and a 2013 update continues to list improved prophylaxis for VTE as a top 10 patient safety strategy to act on now…. The American Public Health Association has stated that the "disconnect between evidence and execution as it relates to DVT prevention amounts to a public health crisis…"

"Various strategies to improve the use of thromboprophylaxis have been demonstrated to be effective, including computerized order sets with electronic alerts, or pre-printed orders and quality improvement in the form of clinician education programs, audit, and feedback, but further efforts are required at improving the translation of data from clinical trials into clinical practice…" Use of PSI 12, and related measures developed by The Joint Commission, encourages providers to adopt the processes or structures of care of the best performing providers, and may empower consumers to select better performing providers or to adhere to recommended prophylactic modalities.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Table 1 is also included in the supplemental materials.

Table 1. Reference Population Rate and Distribution of Hospital Performance for PSI 12 Perioperative Pulmonary Embolism or Deep Vein Thrombosis Rate in 2-year Pooled Data (2011-2013)

Overall Reference Population Rate

Year3	Number of Hospitals	Outcome of Interest (Numerator)1	Population at Risk (Denominator)1	Observed Rate Per 1000 Surgical Discharges1
2011-2012	3,437	46,056	11,638,019	3.9574
2012-2013	3,620	43,301	11,386,129	3.8030

Distribution of Hospital-level Observed Rates in Reference Population

Year3	Number of Hospitals	Rates per 1000 Surgical Discharges (p=percentile)2						
		Mean	SD2	p5	p25	Median	p75	p95
2011-2012	3,437	3.07	3.51	0.00	1.06	2.72	4.37	7.61
2012-2013	3,620	2.98	3.31	0.00	0.87	2.61	4.19	7.38

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2011 - 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov (AHRQ QI Software Version 6.0)

1The observed rate refers to the total rate for all observations included in the reference population data (numerator) divided by the total combined eligible population of all hospitals included in the reference population data (denominator).
2The distribution of hospital rates reports the mean and standard deviation (SD) of the observed rates for all hospitals in the dataset with at least one case in the denominator, as well as the observed rate for hospitals in the 5th, 25th, 50th (median), 75th, and 95th percentile. Standard deviation refers to the spread in observed values in relation to the mean.
3 Reference population is limited to states with present on admission data (POA). Since many states did not report POA data prior to 2011 we have not included testing prior to 2011.

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*
Table 2 is also included in the supplemental materials.

Table 2. Weighted Rates for PSI 12 Perioperative Pulmonary Embolism or Deep Vein Thrombosis Rate (per 1,000 surgical discharges)

Group	PSI 12 v6.0							
	Observed Rate		Risk Adjusted Rate			Standard Error	95% Confidence Interval	Numerator	Denominator
Overall	3.707	3.710	0.025	3.661	3.758	22,235	5,997,737		
Gender									
Male	3.928	3.928	0.037	3.855	4.001	11,032	2,808,568		
Female	3.513	3.518	0.033	3.453	3.583	11,203	3,189,169		
Age									
18-44	2.218	2.217	0.046	2.127	2.308	2,272	1,024,063		
45-64	3.357	3.361	0.038	3.286	3.435	7,761	2,311,921		
65+	4.585	4.588	0.041	4.507	4.669	12,203	2,661,754		
Payer									
Medicare		4.409	4.413	0.039	4.336	4.490	12,475	2,829,236	
Medicaid		3.926	3.922	0.090	3.746	4.098	1,877	478,144	
Private	2.912	2.917	0.037	2.845	2.989	6,163	2,116,711		
Other	2.971	2.970	0.100	2.775	3.165	876	294,985		
Self Pay/Uninsured			3.026	3.022	0.103	2.819	3.224	843	278,661
Race/Ethnicity(1)									
White	3.663	3.674	0.029	3.616	3.731	15,345	4,189,545		
Black	5.000	5.008	0.088	4.835	5.181	3,161	632,244		
Hispanic	2.979	2.983	0.075	2.836	3.130	1,563	524,639		
Asian/Pacific Islander			2.761	2.758	0.147	2.470	3.046	348	125,932
Other	3.461	3.453	0.081	3.295	3.611	1,818	525,377		
Residence									
Non-Metro		2.787	1.974	0.063	1.851	2.098	1,352	485,064	
Metropolitan		3.788	3.796	0.026	3.745	3.847	20,883	5,512,674	

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov (AHRQ QI Software Version 6.0)
1. Hospitals missing race data are excluded. Weighted to approximate national estimates.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
N/A

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, A leading cause of morbidity/mortality, Patient/societal consequences of poor quality
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**

The total number of PSI 12 events per year is shown in Table 1 and in the supplemental materials. Among community hospitals in the AHRQ QI present on admission (POA) Reference Population, derived from the State Inpatient Databases (SID) of the AHRQ Healthcare Cost and Utilization Project (HCUP), the risk-adjusted rate of this indicator was 3.71 per 1,000 eligible patients in 2013. About 20,438 of these adverse events are estimated to have occurred in US community hospitals in 2013. We show substantial variation between 2-year hospital rates for PSI 12, with a coefficient of variation (standard deviation/mean) of 1.14 in 2012-2013. In the peer-reviewed literature, cases from the HCUP Nationwide Inpatient Sample that were flagged by this PSI had 6.6% excess mortality, 5.4 days of excess hospitalization, and $21,709 in excess hospital charges, relative to carefully matched controls that were not flagged (Zhan and Miller,2003). This finding was confirmed in the Veterans Affairs hospital system, where cases that were flagged by this PSI had 6.1% excess mortality, 4.5-5.5 days of excess hospitalization, and $7,205-9,064 in excess hospital costs, relative to carefully matched controls that were not flagged (Rivard et al., 2008). In another study based on HCUP SID from seven states that permit linkage of serial hospitalizations, this indicator was associated with risk ratios of 1.35 for inpatient death, 1.28 for readmission within three months, and 1.25 for readmission within one month (after adjusting for age, gender, payer, comorbidities, specific surgical DRGs, and APR-DRG severity levels) (Friedman et al., 2009). Similarly, in a multivariable analysis of Veterans Health Administration data, hospitalizations with a PSI 12 event were 33% more likely to result in a readmission within 30 days, after adjusting for age, sex, comorbidities, and other PSI events (Rosen et al., 2013).

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2011-2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov (AHRQ QI Software Version 6.0)

Friedman B, Encinosa W, Jiang HJ, Mutter R. Do patient safety events increase readmissions? Med Care 2009; 47(5):583-90.

Rivard PE, Luther SL, Christiansen CL, Zhao S, Loveland S, Elixhauser E, Romano PS, Rosen AK. Using Patient Safety Indicators to estimate the impact of potential adverse events on outcomes. Med Care Res Rev 2008; 65(1):67-87.

Rosen AK, Loveland S, Shin M, et al. Examining the impact of the AHRQ Patient Safety Indicators (PSIs) on the Veterans Health Administration: the case of readmissions. Medical care. 2013;51(1):37-44.

Zhan C, Miller MR. Excess length of stay, charges, and mortality attributable to medical injuries during hospitalization. JAMA 2003;290(14):1868-1874.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

Not applicable

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria 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):*

Surgery : General Surgery

De.6. **Cross Cutting Areas** *(check all the areas that apply):*

Safety : Complications

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.)*

http://1.usa.gov/222tlZu Note: The URL link will be updated for version 6.0 public release found via the module page: http://qualityindicators.ahrq.gov/Modules/psi_resources.aspx

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool

(MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** PSI12_Technical_Specifications_v6.0_160513.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
This revised version will be implemented in forthcoming version 6.0 specifications in 2016. This version (6.0) includes the following changes from the previously-endorsed version (4.4):

- isolated deep vein thrombosis of the calf veins (453.42) is no longer included in the numerator (453.42). This change addresses evidence and concerns from providers that there is substantial inter-hospital variability in the ascertainment and documentation of calf vein thromboses, which are often asymptomatic and have uncertain clinical significance.
- extracorporeal membrane oxygenation (ECMO) procedures are excluded from the denominator (39.65). This change addresses the known very high risk of VTE among patients on ECMO, and the routine use of anticoagulation in this setting.

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)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Discharges, among cases meeting the inclusion and exclusion rules for the denominator, with a secondary ICD-9-CM or ICD-10-CM diagnosis code for proximal deep vein thrombosis or a secondary ICD-9-CM or ICD-10-CM diagnosis code for pulmonary embolism.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
The recommended time period is two years for users with a complete sample of hospital discharges (i.e., "all payer" data). Note that the signal variance parameters in software assume at least a one-year time period. Users may use longer time periods if desired.

S.6. **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, 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.
Please see attached excel file in S.2b. for version 6.0 specifications.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Surgical discharges, for patients ages 18 years and older, with any-listed ICD-9-CM or ICD-10-PCS procedure codes for an operating room procedure. Surgical discharges are defined by specific MS-DRG codes.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
Please see Patient Safety Indicators Appendices in attached excel file in S.2b. for version 6.0 specifications.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Exclude cases:

- with a principal ICD-9-CM or ICD-10-CM diagnosis code (or secondary diagnosis present on admission) for proximal deep vein thrombosis
- with a principal ICD-9-CM or ICD-10-CM diagnosis code (or secondary diagnosis present on admission) for pulmonary embolism
- where a procedure for interruption of vena cava occurs before or on the same day as the first operating room procedure*

- any-listed ICD-9-CM or ICD-10-PCS procedure code for extracorporeal membrane oxygenation (ECMO)
- any-listed ICD-9-CM or ICD-10-CM diagnosis code for acute brain or spinal injury present on admission
- MDC 14 (pregnancy, childbirth, and puerperium)
- with missing gender (SEX=missing), age (AGE=missing), quarter (DQTR=missing), year (YEAR=missing), or principal diagnosis (DX1=missing)

*If day of procedure is not available in the input data file, the rate may be slightly lower than if the information was available.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
Please see attached excel file in S.2b. for version 6.0 specifications.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
Not applicable

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
The predicted value for each case is computed using a hierarchical model (logistic regression with hospital random effect) and covariates for gender, age (in 5-year age groups, except for the youngest age range), Modified Diagnosis Related Groups, which are the base MS DRGs without any distinction for "comorbidity and complications" (CC/MCC), AHRQ Comorbidity Index, Major Diagnosis Categories (MDC) based on the principal diagnosis, and transfer in from another acute care hospital. A parsimonious model was identified using a backward stepwise selection procedure with bootstrapping. The expected rate is computed as the sum of the predicted value for each case divided by the number of cases for the unit of analysis of interest (i.e., hospital). The risk adjusted rate is computed using indirect standardization as the observed rate divided by the expected rate, multiplied by the reference population rate.

Additional information on methodology can be found in the Empirical Methods document on the AHRQ Quality Indicator website (www.qualityindicators.ahrq.gov). The Empirical Methods are also attached in the supplemental materials.

The specific covariates for this measure are as follows:

PARAMETER	LABEL
Intercept	Intercept
Sex | Age Demographics	
M_AgeCat_1	Male | Age 18 - 29
M_AgeCat_2	Male | Age 30 - 34
M_AgeCat_3	Male | Age 35 - 39
M_AgeCat_4	Male | Age 40 - 44
M_AgeCat_5	Male | Age 45 - 49
M_AgeCat_6	Male | Age 50 - 54
M_AgeCat_7	Male | Age 55 - 59
M_AgeCat_8	Male | Age 60 - 64
M_AgeCat_9	Male | Age 65 - 69
M_AgeCat_10	Male | Age 70 - 74
M_AgeCat_11	Male | Age 75 - 79
M_AgeCat_12	Male | Age 80 - 84
M_AgeCat_13	Male | Age 85 - 89
M_AgeCat_14	Male | Age >=90
F_AgeCat_1	Female | Age 18 - 29
F_AgeCat_2	Female | Age 30 - 34

F_AgeCat_3	Female | Age 35 - 39
F_AgeCat_4	Female | Age 40 - 44
F_AgeCat_5	Female | Age 45 - 49
F_AgeCat_6	Female | Age 50 - 54
F_AgeCat_7	Female | Age 55 - 59
F_AgeCat_8	Female | Age 60 - 64
F_AgeCat_9	Female | Age 65 - 69
F_AgeCat_10	Female | Age 70 - 74
F_AgeCat_11	Female | Age 75 - 79
F_AgeCat_12	Female | Age 80 - 84
F_AgeCat_13	Female | Age 85 - 89
F_AgeCat_14	Female | Age >=90
Origin	
TRNSFER	Transfer from another facility
Comorbidities	
ANEMDEF	Deficiency Anemias
BLDLOSS	Chronic blood loss anemia
CHF	Congestive heart failure
COAG	Coagulopathy
DEPRESS	Depression
DM	Diabetes w/o chronic complications
DMCX	Diabetes w/ chronic complications
HTN_C	Hypertension, Complicated
HYPOTHY	Hypothyroidism
IMMUNE	Immune disorders
LIVER	Liver disease
LYMPH	Lymphoma
LYTES	Fluid and electrolyte disorders
METS	Metastatic cancer
OBESE	Obesity
PARA	Paralysis
PSYCH	Psychoses
PULMCIRC	Pulmonary circulation disease
RENLFAIL	Renal failure
TUMOR	Solid tumor w/out metastasis
WGHTLOSS	Weight loss
Major Diagnostic Categories (MDC)	
MDC_1	MDC 1: Nervous System
MDC_3	MDC 3: Ear Nose Mouth And Throat
MDC_4	MDC 4: Respiratory System
MDC_5	MDC 5: Circulatory System
MDC_6	MDC 6: Digestive System
MDC_7	MDC 7: Hepatobiliary System And Pancreas
MDC_8	MDC 8: Musculoskeletal And Connective
MDC_9	MDC 9: Skin Subcutaneous And Breast
MDC_10	MDC 10: Endocrine Nutritional And Metabolic
MDC_11	MDC 11: Kidney And Urinary Tract
MDC_13	MDC 13: Female Reproductive System
MDC_16	MDC 16: Blood and Immunological
MDC_18	MDC 18: Infectious and Parasitic
MDC_20	MDC 20: Alcohol/Drug Disorders
MDC_21	MDC 21: Injuries Poison And Toxic
MDC_22	MDC 22: Burns
MDC_23	MDC 23: Factors Influencing Health
Modified Diagnostic Related Groups (MDRG)	
mdrg_1001	Adrenal & pituitary procedures
mdrg_1002	Amputation of lower limb for endocrine

mdrg_1003	O.R. procedures for obesity
mdrg_1004	Skin grafts & wound debridement for endoc
mdrg_1005	Thyroid parathyroid & thyroglossal procedures
mdrg_1006	Other endocrine nutritional & metabolic procedures
mdrg_102	Craniotomy w major dev impl/acute complex CNS
mdrg_103	Craniotomy & endovascular intracranial procedures
mdrg_104	Spinal procedures
mdrg_105	Ventricular shunt procedures
mdrg_106	Carotid artery stent procedure
mdrg_107	Extracranial procedures
mdrg_108	Peripheral & cranial nerve & other nervous system procedures
mdrg_1101	Kidney transplant
mdrg_1102	Major bladder procedures
mdrg_1103	Kidney & ureter procedures for neoplasm
mdrg_1104	Kidney & ureter procedures for non-neoplasm
mdrg_1105	Minor bladder procedures
mdrg_1106	Prostatectomy
mdrg_1107	Transurethral procedures
mdrg_1108	Urethral procedures
mdrg_1109	Other kidney & urinary tract procedures
mdrg_1201	Major male pelvic procedures
mdrg_1202	Penis procedures
mdrg_1203	Testes procedures
mdrg_1204	Transurethral prostatectomy
mdrg_1301	Pelvic evisceration - radical hysterectomy
mdrg_1302	Uterine & adnexa procedures ovarian or adnexal malignancy
mdrg_1303	Uterine adnexa procedures non-ovarian/adnexal malignancy
mdrg_1304	Uterine & adnexa procedures for non-malignancy
mdrg_1305	D&C conization laparoscopy & tubal interruption
mdrg_1306	Vagina cervix & vulva procedures
mdrg_1307	Female reproductive system reconstructive
mdrg_1308	Other female reproductive system procedures
mdrg_1601	Splenectomy
mdrg_1602	Other O.R. procedures of the blood & blood forming
mdrg_1707	Lymphoma & leukemia
mdrg_1708	Lymphoma & non-acute leukemia
mdrg_1801	Infectious & parasitic diseases w procedure
mdrg_1802	Postoperative or post-traumatic infections
mdrg_2101	Wound debridements for injuries
mdrg_2102	Skin grafts for injuries
mdrg_2103	Hand procedures for injuries
mdrg_2104	Other O.R. procedures for injuries
mdrg_2201	Full thickness burn w skin graft or inhalation injury
mdrg_2210	Extensive burns or full thickness burns
mdrg_2301	O.R. procedures w diagnoses of other contact
mdrg_2407	Limb reattachment hip & femur procedures
mdrg_2408	Other O.R. procedures for multiple sig trauma
mdrg_301	Acute major eye infections
mdrg_302	Other ear nose mouth & throat O.R. procedures
mdrg_304	Mouth procedures
mdrg_305	Salivary gland procedures
mdrg_401	Major chest procedures
mdrg_402	Other respiratory system O.R. procedures
mdrg_502	Percutaneous cardiovascular procedures w non-drug-eluting stent
mdrg_503	Cardiac valve & other major cardiothoracic procedures
mdrg_504	Cardiac defibrillator implant
mdrg_505	Other cardiothoracic procedures

mdrg_506	Coronary bypass w PTCA
mdrg_507	Coronary bypass w cardiac catheterization
mdrg_509	Amputation for circulatory sys disorders
mdrg_510	Permanent cardiac pacemaker implant
mdrg_511	Percutaneous cardiovascular procedures w drug-eluting stent
mdrg_513	Percutaneous cardiovascular procedures w/o coronary artery stent
mdrg_514	Other vascular procedures
mdrg_515	Upper limb & toe amputation
mdrg_516	Cardiac pacemaker device replacement
mdrg_517	Cardiac pacemaker revision
mdrg_519	Other circulatory system O.R. procedures
mdrg_601	Stomach esophageal & duodenal procedures
mdrg_602	Major small & large bowel procedures
mdrg_603	Rectal resection
mdrg_604	Peritoneal adhesiolysis
mdrg_605	Appendectomy w complicated principal diagnosis
mdrg_606	Appendectomy w/o complicated principal diagnosis
mdrg_607	Minor small & large bowel procedures
mdrg_608	Anal & stomal procedures
mdrg_609	Inguinal & femoral hernia procedures
mdrg_610	Hernia procedures except inguinal & femoral
mdrg_611	Other digestive system O.R. procedures
mdrg_701	Pancreas liver & shunt procedures
mdrg_702	Biliary tract procedures except only cholecystectomy
mdrg_703	Cholecystectomy w common duct exploration
mdrg_704	Cholecystectomy except by laparoscope
mdrg_705	Laparoscopic cholecystectomy
mdrg_706	Hepatobiliary diagnostic procedures
mdrg_707	Other hepatobiliary or pancreas procedures
mdrg_7701	Heart transplant or implant heart assist system
mdrg_7702	Liver transplant
mdrg_7703	Lung transplant
mdrg_801	Combined anterior/posterior spinal fusion
mdrg_802	Spinal fusion except cervical w spinal curvature/malignancy/infection
mdrg_803	Spinal fusion except cervical
mdrg_804	Bilateral or multiple major joint procedures
mdrg_805	Wnd debridement & skin graft excision hand for musculoskeletal
mdrg_806	Revision of hip or knee replacement
mdrg_807	Major joint replacement or reattachment
mdrg_808	Cervical spinal fusion
mdrg_809	Amputation for musculoskeletal system
mdrg_810	Biopsies of musculoskeletal system
mdrg_811	Hip & femur procedures except major joint
mdrg_812	Major joint & limb reattachment
mdrg_813	Knee procedures w principal diagnosis of infection
mdrg_814	Knee procedures w/o principal diagnosis of infection
mdrg_815	Back & neck procedures exc spinal fusion
mdrg_816	Lower extremity & humerus procedures
mdrg_817	Local excision & removal internal fixation devices
mdrg_818	Local excision & removal internal fixation devices
mdrg_819	Soft tissue procedures
mdrg_820	Foot procedures
mdrg_821	Major thumb or joint procedures
mdrg_822	Major shoulder or elbow joint procedures
mdrg_824	Shoulder elbow or forearm procedures
mdrg_825	Hand or wrist procedures
mdrg_826	Other musculoskeletal system & connective tissue procedures

mdrg_8899	Non-Extensive O.R. Procedures Unrelated to PDX
mdrg_901	Skin graft &/or debridement for skin ulcer or cellulitis
mdrg_902	Skin graft &/or debridement except for skin ulcer
mdrg_903	Other skin subcutaneous tissue & breast procedures
mdrg_904	Mastectomy for malignancy
mdrg_905	Breast biopsy local excision

c-statistic = .751

Source: http://www.qualityindicators.ahrq.gov/Modules/psi_resources.aspx
Parameter estimates are also included with the Technical Specifications attached in section S.2b

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Available in attached Excel or csv file at S.2b

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
Available in attached Excel file at S.2b

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **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.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
The observed rate is the number of discharge records where the patient experienced the PSI adverse event divided by the number of discharge records at risk for the event. The expected rate is a comparative rate that incorporates information about a reference population that is not part of the user's input dataset – what rate would be observed if the expected level of care observed in the reference population and estimated with risk adjustment regression models, were applied to the mix of patients with demographic and comorbidity distributions observed in the user's dataset. The expected rate is calculated only for risk-adjusted indicators.

The expected rate is estimated for each person using a generalized estimating equations (GEE) approach to account for correlation at the hospital or provider level.

The risk-adjusted rate is a comparative rate that also incorporates information about a reference population that is not part of the input dataset – what rate would be observed if the level of care observed in the user's dataset were applied to a mix of patients with demographics and comorbidities distributed like the reference population? The risk adjusted rate is calculated using the indirect method as observed rate divided by expected rate multiplied by the reference population rate. The smoothed rate is the weighted average of the risk-adjusted rate from the user's input dataset and the rate observed in the reference population; the smoothed rate is calculated with a shrinkage estimator to result in a rate near that from the user's dataset if the provider's rate is estimated in a stable fashion with minimal noise, or to result in a rate near that of the reference population if the variance of the estimated rate from the input dataset is large compared with the hospital-to-hospital variance estimated from the reference population. Thus, the smoothed rate is a weighted average of the risk-adjusted rate and the reference population rate, where the weight is the signal-to-noise ratio. In practice, the smoothed rate brings rates toward the mean, and tends to do this more so for outliers (such as rural hospitals).
For additional information, please see the supplemental materials for the AHRQ QI Empirical Methods.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Not applicable

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not applicable

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Exclude cases with missing gender (SEX=missing), age (AGE=missing), quarter (DQTR=missing), year (YEAR=missing), or principal diagnosis (DX1=missing). Missingness on these variables, in aggregate, almost never exceeds 1% of eligible records.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
While the measure is tested and specified using data from the Healthcare Cost and Utilization Project (HCUP) (see section 1.1 and 1.2 of the measure testing form), the measure specifications and software are specified to be used with any ICD-9-CM or ICD-10-CM/PCS coded administrative billing/claims/discharge dataset.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
Not applicable

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
PSI12_Measure_Testing_Form_160513_v5.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 0450
Measure Title: Perioperative Pulmonary Embolism or Deep Vein Thrombosis Rate (PSI 12)
Date of Submission: 5/13/2016
Type of Measure:

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

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decision making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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.23*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ administrative claims	☒ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

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

All analyses were completed using data from the Healthcare Cost and Utilization Project (HCUP) State Inpatient Databases (SID), 2011-2013. HCUP is a family of health care databases and related software tools and products

developed through a Federal-State-Industry partnership and sponsored by the Agency for Healthcare Research and Quality (AHRQ).[1] HCUP databases bring together the data collection efforts of State data organizations, hospital associations, private data organizations, and the Federal government to create a national information resource of encounter-level health care data. The HCUP SID contain the universe of the inpatient discharge abstracts in participating States, translated into a uniform format to facilitate multi-State comparisons and analyses. All states provide data for community hospitals and together, the SID encompasses about 97 percent of all U.S. community hospital discharges. For the analyses presented here, we use 34 states representing about 89 percent of the U.S. community hospital discharges, for a total of about 30 million hospital discharges from community hospitals. As defined by the American Hospital Association, community hospitals are all non-Federal, short-term, general or other specialty hospitals, excluding hospital units of institutions. Included among community hospitals are public and academic medical centers, specialty hospitals such as obstetrics–gynecology, ear–nose–throat, orthopedic and pediatric institutions. Short-stay rehabilitation, long-term acute care hospitals are excluded from the data used for the reported analyses.

Each of the 34 states included in the dataset report information about whether a diagnosis was present on admission (POA) and information on the timing of procedures during the hospitalization. POA data[2] is important to distinguish complications that occur in-hospital from diagnoses that existed prior to hospitalization. Edit checks on POA were developed using a separate analysis of HCUP databases that examined POA coding in the 2013 SID at hospitals that were required to report POA to CMS. The edits identify general patterns of suspect reporting of POA. The edits do not evaluate whether a valid POA value (e.g., Y or N) is appropriate for the specific diagnosis. There are three hospital-level edit checks:

1. Indication that a hospital has POA reported as Y on all diagnoses on all discharges
2. Indication that a hospital has POA reported as missing on all non-Medicare discharges
3. Indication that a hospital reported POA as missing on all nonexempt diagnoses for 15 percent or more of discharges. The cut-point of 15 percent was determined by 2 times the standard deviation plus the mean of the percentage for hospitals required to report POA to CMS.

Hospitals that failed any of the edit checks were excluded from the dataset.
The SID data elements include International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) coded principal and secondary diagnoses and procedures, additional detailed clinical and service information based on revenue codes, admission source and discharge status, patient demographics, expected payment source (Medicare, Medicaid, private insurance as well as the uninsured), total charges and length of stay (www.hcup-us.ahrq.gov).

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

HCUP data included 2011-2013, for most tests we combine hospital data for 2 years prior to calculating rates and testing the measure. This is termed “in 2-year pooled data” in the results below.

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

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.26***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency

[1]HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2011-2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)
[2] Present-on -Admission was added as a data element to the uniform bill form (UB-04) effective October 1, 2007, and hospitals incurred a payment penalty for not including POA on Medicare records beginning October 1, 2008. Each of the several diagnoses in a discharge record can be flagged as “present at the time the order for inpatient admission occurs” or not (see http://www.cdc.gov/nchs/icd/icd9cm_addenda_guidelines.htm).

☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

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

Table 1. Reference Population Rate and Distribution of Hospital Performance for PSI 12 Perioperative Pulmonary Embolism or Deep Vein Thrombosis Rate in 2-year Pooled Data (2011-2013)

Overall Reference Population Rate				
Year[3]	Number of Hospitals	Outcome of Interest (Numerator)[1]	Population at Risk (Denominator)[1]	Observed Rate Per 1000 Surgical Discharges[1]
2011-2012	3,437	46,056	11,638,019	3.9574
2012-2013	3,620	43,301	11,386,129	3.8030

Distribution of Hospital-level Observed Rates in Reference Population								
Year[3]	Number of Hospitals	Rates per 1000 Surgical Discharges (p=percentile)[2]						
		Mean	SD[2]	p5	p25	Median	p75	p95
2011-2012	3,437	3.07	3.51	0.00	1.06	2.72	4.37	7.61
2012-2013	3,620	2.98	3.31	0.00	0.87	2.61	4.19	7.38

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2011 - 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov (AHRQ QI Software Version 6.0)

[1]The observed rate refers to the total rate for all observations included in the reference population data (numerator) divided by the total combined eligible population of all hospitals included in the reference population data (denominator).

[2]The distribution of hospital rates reports the mean and standard deviation (SD) of the observed rates for all hospitals in the dataset with at least one case in the denominator, as well as the observed rate for hospitals in the 5th, 25th, 50th (median), 75th, and 95th percentile. Standard deviation refers to the spread in observed values in relation to the mean.

[3] Reference population is limited to states with present on admission data (POA). Since many states did not report POA data prior to 2011 we have not included testing prior to 2011.

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

See 1.5 (Table 1)

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.

Some tests require comparisons across two or three years of data (2011-2013). When no comparisons are required for the test, typically 2013 data are used. Some validity testing uses only 2011 or 2012 data.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Age and sex were the only patient-level sociodemographic variables that were available and analyzed in the data used for measure development and testing. Many of the HCUP SID include race/ethnicity, and all of the HCUP SID include the primary expected source of payment, and zip code of residence, which could be used to capture socioeconomic characteristics at an ecological (community) level. While these variables were used to assess disparities at the national level, these variables were not used in the current risk adjustment model, based on our conceptual description (i.e., logical rationale or theory informed by literature and content experts) of the causal pathway between these factors, patient clinical factors, quality of care, and outcome, described in Section 2b4.3 below.

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*)
Signal-to-Noise. The signal-to-noise ratio is a measure of reliability that is calculated at the hospital level and then summarized across the entire population of US hospitals. It compares the degree to which risk adjusted rates differ across hospitals (the signal) to the degree of precision of the rates within hospitals (the noise). This metric is a stringent measure of reliability that takes into account the observed distribution of risk adjusted rates within a reference population. An indicator with a low signal-to-noise ratio may not be able to distinguish differences in performance among hospitals, or may identify differences inconsistently within the same time period. An indicator with a high signal-to-noise ratio will be more likely to consistently distinguish performance differences among hospitals (e.g. one hospital performs better than others).

The signal-to-noise ratio is estimated for each hospital. The overall signal-to-noise estimate is an average of hospital-level signal to noise ratios weighted by a value of one divided by the signal plus the hospital's noise for PSI 12. Hospitals with smaller denominators (the number of patients at risk) will have lower weight, and less influence on the overall signal-to-noise ratio, because of higher noise. Weighting reduces the influence of hospitals that have less reliable rates due to very small denominators (the number of patients at risk) on the overall signal-to-noise ratio estimate.

Because the signal-to-noise ratio quantifies the ability to consistently discriminate one hospital's performance from the other hospitals in the population, it is sensitive to the distribution of hospital sizes as well as the distribution of risk-adjusted rates in the reference population. If the hospitals in a population all have performance in a narrow range (low signal), it is more difficult to reliably distinguish among hospitals' performance than when hospital performance is spread out over a much wider range (high signal). For example, if all hospitals have nearly perfect performance, it will be impossible to distinguish among them. As a consequence, if the distribution of hospital rates changes over time, the signal-to-noise ratio will also change.

There is no universally accepted threshold of "adequate" signal to noise ratio. Different methods of calculating reliability and signal-to-noise (e.g., split sample or test-retest reliability of the data, different methods of calculating the hospital signal-to-noise ratio) result in different distributions of reliability scores. In addition, "adequate" depends on the specific application and judgment of the user. For instance, if a complication such as mortality is very important (e.g. leads to great harm to the patient) a lower reliability may be acceptable. However, the AHRQ QI program generally considers ratios between 0.4 – 0.8 as acceptable. It is rare to achieve reliability above 0.8, using hospital signal-to-noise ratios as an indicator of reliability. To account for the uncertainty (noise) in a hospital's performance due to low volume, a longer period of data can be used or smoothed rates can be calculated.

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*)
Table 2 shows the most recent reliability testing for PSI 12.

Table 2. Signal-to-Noise Ratio by Hospital Size Decile, PSI 12 Perioperative Pulmonary Embolism or Deep Vein Thrombosis Rate Using in 2-year Pooled Data (2012-2013)

Hospital Size Decile	Number of Hospitals	Avg. Number of Discharges per Hospital in Decile	Avg. Signal-to-Noise Ratio for Hospitals in Decile
1 (smallest)	357	45.3	0.0561

2	358	228.5	0.1595
3	357	541.9	0.2895
4	358	998.3	0.4197
5	357	1,608.1	0.5362
6	358	2,372.7	0.6275
7	358	3,302.9	0.7024
8	357	4,548.3	0.7675
9	358	6,561.0	0.8298
10 (largest)	357	11,648.9	0.8975
Overall	3,575	3,184.9	0.7359

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2012 - 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov (AHRQ QI Software Version 6.0)

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?*)

Signal-to-noise ratios were smaller for hospitals with fewer than approximately 100 qualifying discharges per year (average signal-to-noise ratio less than 0.42). Smoothed rates, which are recommended for all hospitals (and are implemented in the AHRQ software), address reliability concerns particularly for small hospitals. Hospitals with more than 1000 qualifying discharges on average have risk adjusted rates with moderate to high reliability (average signal-to-noise ratio of 0.42 to 0.90). Overall, the signal to noise ratio for this indicator is good with an overall signal-to-noise ratio of 0.74.

2b2. VALIDITY TESTING

2b2.1. 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*)

2b2.2. 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)*

Critical Data Elements

Several studies assessing the criterion validity of PSI 12, using medical record abstraction by trained nurses as a gold standard, have been published in the peer review literature. We summarize the most relevant literature in this form, and provide a full evidence summary in the attached Evidence Form.

Performance Measure Score

Empirical Validity Analyses

We conducted analyses focused on specific critiques of PSI 12 using the AHRQ QI Reference Population described above. One such critique is that PSI 12 should include only elective admissions, which typically have the index procedures on day 0 or 1 of the hospitalization. For this analysis we stratified the PSI 12 events in the 2012 AHRQ QI Reference Population by preoperative length of stay (days before the index procedure). We then calculated the PSI 12 rate for each stratum.

Systematic Assessment of Face Validity

We utilized a structured panel review to evaluate face validity (from a clinical perspective) of the Patient Safety Indicators. The panels were convened in 2002. It is anticipated that the results of face validity review would be similar if panels were convened in more recent years, given that the clinical characteristics of these events, treatment and prevention approaches, and sequelae have not changed substantially since 2002. The clinical panel review process was based on the RAND appropriateness method, a modified Delphi process also known as a nominal group technique.

Twenty-one professional clinical organizations were invited to submit nominations. These organizations were selected based on the applicability of the specialty or subspecialty to potential Patient Safety Indicators. Clinical areas represented by the panels included internal medicine, cardiology, radiology, geriatrics, surgical and critical care nursing, anesthesiology, pharmacy, inpatient medicine and surgery (including thoracic, neurology, orthopedic, colorectal, urology, spine, and transplant surgical subspecialties). For assignments to each panel, a list of applicable specialties was identified for the indicators to be evaluated by that panel. Panelists were selected so that each panel had diverse membership in terms of practice characteristics and setting. For PSI 12, 7 members of a multispecialty panel and 6 members of a surgical subspecialty panel completed the evaluation in full. Additional details of panel composition are available online at http://archive.ahrq.gov/clinic/tp/hospdatp.htm.

Panelists completed a 10-item questionnaire, tailored to each specific indicator. Following the initial rating of the indicators, panelists participated in a moderated 90-minute conference call, where opinions about the indicators were discussed. The panelists then completed the same 10-item questionnaire again, and submitted their final ratings. Ratings were summarized in accordance with the RAND Appropriateness Method.[3]

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Critical Data Elements

Several studies have evaluated the criterion validity of administrative data in detecting postoperative PE/DVT. Most examined data prior to 2010, when significant changes to the specifications were made to incorporate changes in the ICD-9-CM codes for DVT (described below). These studies are described in full in the PSI 12 Evidence Form included in this submission packet, but in summary, they reported positive predictive values (PPVs) of 43% (95% CI, 34-53%) in VA hospitals[4,5], 44% (95% CI, 37-51%) in academic medical centers, and 48% (95% CI, 42-52%) in a national sample of volunteer hospitals.[6] False-negative errors were extremely rare, with an estimated sensitivity in the academic sample of 100% for identifying acute lower extremity or pelvic VTE and 95.5% for identifying any acute venous thrombosis. These studies were recently meta-analyzed by Winters et al.,[6] who reported a pooled PPV estimate of 63% (95% CI, 44-83%), but this estimate is incorrect because the authors double-counted the same VA study,**Error! Bookmark not defined.**,**Error! Bookmark not defined.** failed to account for POA reporting, and failed to include more recent studies cited below.

In a combined analysis of 573 PSI-flagged cases from the second and third of these samples, 74 (12.9%) had a documented prior/chronic VTE, which was presumably present at admission, 73 (12.7%) had an acute VTE before the operation, 19 (3.3%) had an acute VTE of undetermined timing, 83 (14.5%) had acute upper extremity thrombosis), 34 (5.9%) had superficial vein thrombosis, and 12 (2.1%) had thrombosis of unknown site. Only 48 (8.4%) of flagged cases had no mention of VTE in the abstracted record. It should be noted that preoperative but hospital-acquired VTEs were classified as false positives in these studies, even though AHRQ now argues that these should be considered as true positives (because many are known to be related to delays in surgery or ineffective preoperative prophylaxis).

---

[3] McDonald KM, Romano PS, Geppert J, Davies SM, Duncan BW, Shojania KG. Measures of Patient Safety Based on Hospital Administrative Data: The Patient Safety Indicators. Technical Review Number 5. Rockville, MD: Agency for Healthcare Research and Quality, 2002

[4] Kaafarani HM, Borzecki AM, Itani KM et al. Validity of selected Patient Safety Indicators: opportunities and concerns. J Am Coll Surg. 2011 Jun;212(6):924-34.

[5] Rosen AK, Itani KMF, Cevasco M, et al. Validating the Patient Safety Indicators in the Veterans Health Administration. Med Care. 2012;50:74-85.

[6] Winters BD, Bharmal A, Wilson RF, et al. Validity of the Agency for Health Care Research and Quality Patient Safety Indicators and the Centers for Medicare and Medicaid Hospital-acquired Conditions: A Systematic Review and Meta-Analysis. *Medical Care.* 2016.

After reviewing these data with various stakeholders, AHRQ concluded that PSI 12 captured upper extremity and superficial thromboses because the existing ICD-9-CM diagnosis codes lacked specificity; codes for these diagnoses were available under the "thrombophlebitis" heading (451.xx), but not under the much more commonly used "thrombosis" heading (453.xx). In addition, coders reported confusion about how to code chronic thromboses that are diagnosed after admission. Based on these findings and other studies in the peer-reviewed literature, AHRQ proposed and the ICD-9-CM Coordination and Maintenance Committee implemented an entirely new set of ICD-9-CM codes for superficial, upper extremity, and chronic venous thromboses. These codes are now excluded from the definition of PSI 12, which prompted AHRQ to reexamine the PPV of this indicator.

Sadeghi et al. used two sources of data from 2009-2010 to generate updated estimates: (1) a retrospective case-control study of risk factors for acute VTE after TKA in 15 teaching hospitals; (2) a chart abstraction data by 7 volunteer hospitals participating in AHRQ's Validation Pilot Project. In the UHC sample, the PPV of PSI 12 was 99% (125/126) and the NPV was 99.4% (460/463). In the AHRQ sample, the overall PPV was 81% (126/156) and most false positives were attributable to incomplete reporting of POA status (which now results in exclusion from the AHRQ reference population).[7] Ramanathan et al. similarly reported a PPV of 88% (95% CI: 80-93%) from FY 2012 data at a single academic medical center in Virginia.[8] Finally, Quan et al. randomly sampled patients in 2007 and 2008 with PSI events from 3 Calgary hospitals; the PPV for PSI12 was 90% (95% CI 67% to 99%).[9]

At least four other studies assessed PSI 12 relative to clinical registries that track postoperative VTE, such as the National Surgical Quality Improvement Program (NSQuIP) and the Veterans Health Administration Surgical Quality Improvement Program (VASQuIP). However, the definition of postoperative VTE in NSQuIP and VASQuIP is based on a positive (or "high probability") imaging test with "initiation of anticoagulation therapy" and/or "placement of mechanical interruption" in the inferior vena cava, whereas the AHRQ definition is based on a physician-documented diagnosis (without regard to how the condition was treated). After linking 55,752 hospitalizations in the 2001 VA inpatient administrative data with the corresponding VASQuIP records, Romano et al. reported that the version 3.0 specification of PSI 12 had a sensitivity of 56% (95% CI, 50-63%) and a PPV of 22% (95% CI, 19-25%).[10] Mull et al. replicated this study with version 4.1 PSI software and 268,771 records from 2003-2007, and reported that PSI 12 had a sensitivity of 65% (95% CI, 63-67%) and PPV of 31% (95% CI, 30-33%).[11] However, when these authors reviewed 20 discrepancies manually (i.e., PSI 12 positive, VASQuIP negative), 14 were actually true positives by PSI 11 and false negatives in VASQuIP. This finding suggests that the actual PPV of PSI 12 is substantially higher than 31%; in other words, VASQuIP cannot be regarded as a gold standard.

Empirical Validity Analyses

Table 3 shows the rates of PSI 12 events (v5.0) by day of index procedure. Index procedures that occur >2 days after admission may represent non-elective operations and suggest the possibility that the numerator event may have occurred before surgery (but after admission). Long preoperative delays may be a mutable process of care.

---

[7] Sadeghi B, White RH, Maynard G, et al. Improved Coding of Postoperative Deep Vein Thrombosis and Pulmonary Embolism in Administrative Data (AHRQ Patient Safety Indicator 12) After Introduction of New ICD-9-CM Diagnosis Codes. *Medical care.* 2015; 53(5):e37-40.

[8] Ramanathan R, Leavell P, Wolfe LG, Duane TM. Agency for Healthcare Research and Quality patient safety indicators and mortality in surgical patients. *The American surgeon.* 2014;80(8):801-804.

[9] Quan H, Eastwood C, Cunningham CT, et al. Validity of AHRQ patient safety indicators derived from ICD-10 hospital discharge abstract data (chart review study). *BMJ open.* 2013;3(10):e003716.

[10] Romano PS, Mull HJ, Rivard PE, et al. Validity of selected AHRQ patient safety indicators based on VA National Surgical Quality Improvement Program data. Health Serv Res. 2009 Feb;44(1):182–204

[11] Mull HJ, Borzecki AM, Loveland S, et al. Detecting adverse events in surgery: comparing events detected by the Veterans Health Administration Surgical Quality Improvement Program and the Patient Safety Indicators. *American journal of surgery.* 2014;207(4):584-595.

Table 3. Observed Rates of PSI 12 events (Perioperative Pulmonary Embolism or Deep Vein Thrombosis) by Day of Index Procedure

Day of index procedure	2011			2012		
	Number of surgical discharges	Number of events	Observed Rate (per 1,000 surgical discharges)	Number of surgical discharges	Number of events	Observed Rate (per 1,000 surgical discharges)
Any	5,922,646	30,894	0.0052	6,033,676	30,478	0.0051
Day 0	3,934,371	15,623	0.0040	3,995,573	15,581	0.0039
Day 1	790,810	4,802	0.0061	814,851	4,570	0.0056
Day 2	364,930	2,281	0.0063	374,115	2,310	0.0062
Day 3	226,445	1,527	0.0067	231,646	1,524	0.0066
Day 4	528,242	6,348	0.0120	530,954	6,140	0.0116
Day >4	606090	6661	0.0110	617491	6493	0.0105

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2011 - 2012. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov (AHRQ QI Software Version 6.0)

Systematic Assessment of Face Validity

The multi-specialty Panel and Surgical Panel both rated the indicator as acceptable on overall usefulness as an indicator of potentially preventable complications of care.

Table 4. Clinician Panel Evaluations of the Face Validity of PSI 12[4]

Multi-specialty Panel (MSP) Evaluation			Surgical Panel (SP) Evaluation		
Overall Rating[1]	Agreement[2]	Acceptability[3]	Overall Rating[1]	Agreement[2]	Acceptability[3]
7	Indeterminate	Acceptable (-)	7	Agreement	Acceptable

[1]Median panel overall rating of the indicator on a scale from 1 to 9, with the higher rating indicating better measurement
[2]Level of agreement, where "agreement" corresponds to little dispersion of opinion, "indeterminate" means that the opinion ranged but did not reach the point of clear "disagreement", the final category where there were panelists with diametrically different opinions
[3]"Acceptable" indicates that the indicator was rated as useful by almost all panelists. "Acceptable (-)" indicates that the indicator was rated as useful by most panelists, although a few rated it as less useful (but not as poor). "Unclear" indicates that panelists rated the usefulness of the indicator as moderate. For further details of methods, see http://archive.ahrq.gov/clinic/tp/hospdatp.htm
[4]PSI 12 was evaluated under a previous name (i.e. Postoperative Pulmonary Embolism or Deep Vein Thrombosis Rate).

2b2.4. 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?*)

Although studies in the peer-review literature reported moderate PPV prior to 2010, changes to the ICD-9-CM codes for VTE events and subsequent modification of the PSI 12 algorithm have improved the performance of the measure. More recent studies have estimated that these changes raised the PPV from 81% to 99%, depending on the clinical setting. The false negative rate appears to be very low for this measure.

In our analysis of PSI 12 rates stratified by preoperative stay, patients with long preoperative delays (4+ days) had 3 times higher risk of a postoperative VTE than patients who went to surgery in a more timely manner. The results suggest that reducing long preoperative delays may improve VTE rates. However, some of these preoperative delays may have been outside the provider's control, if these patients were too sick to undergo surgery before day 4 or did not require surgery until then.

In our clinical panel review, the indicator had high face validity for use in quality improvement and hospital comparative assessments.

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

2b3.1. 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*)

Empirical Evaluation of Exclusions: Using the 2013 data from 34 states, we examined the percent of potential denominator cases excluded by each criterion as listed in the measure specifications.

2b3.2. 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 5 shows the results of the most recent exclusions analysis.

Table 5. Number and Percent of Discharges Excluded, by Denominator Exclusion Criteria, PSI 12 Perioperative Pulmonary Embolism or Deep Vein Thrombosis Rate [1]

PSI 12	Denominator			Potential Numerator[2]		
Exclusion Name	**Exclusion Count**	**After Exclusions**	**% Change**	**Exclusion Count**	**After Exclusions**	**% Change**
No Exclusions applied	**-**	**6,671,854**	**-**	**-**	**45,773**	**-**
Exclude Principal Diagnosis of DVT or PE	27,465	6,644,389	0.41%	22,443	23,330	49%
Exclude if interruption of vena cava occurs before or on the same day as the first OR procedure	15,229	6,656,625	0.2%	6,022	39,751	13.2%
Exclude MDC 14	1,050,160	5,621,694	15.7%	432	45,341	0.9%
Exclude any diagnosis POA of acute brain and or spinal injury	75,374	6,596,480	1.1%	2,486	43,287	5.4%
All exclusions applied	1,150,550	5,512,304	17.2%	25,335	20,438	55%

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov (AHRQ QI Software Version 6.0)

[1]This indicator does not have numerator exclusion criteria.

[2]Potential numerator cases are those that would have qualified for the numerator if not for a particular denominator exclusion criterion.

2b3.3. 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 denominator excludes records with a principal diagnosis of DVT or PE, or a secondary diagnosis of DVT or PE reported as present on admission. In each of these cases, the precipitating events happened before the targeted hospitalization, and are thus unlikely to reflect quality of care during that hospitalization.

The other denominator exclusions are intended to reduce the number of flagged cases in which the diagnosis was non-preventable. For example, patients who had a procedure for interruption of vena cava before or on the same day as the first operating room procedure are excluded because this procedure is only indicated for patients who cannot receive or have already failed conventional pharmacologic prophylaxis; it may reduce the risk of PE at the expense of a higher risk of DVT distal to the occlusive device. Patients receiving extracorporeal membrane oxygenation (ECMO) are excluded because of the high associated risk of VTE despite routine anticoagulation. Cases with any acute brain and/or spinal injuries present on admission are excluded because prophylaxis may place such patients at risk for bleeding, and bleeding in or around the brain or spinal cord may have severe consequences. Women admitted for conditions related

to pregnancy, labor and delivery (Major Diagnosis Category 14, MDC 14) are excluded because this complication is extremely rare in that clinical setting and different ICD-9-CM diagnosis codes from the pregnancy chapter would apply. Empirical analyses support these exclusions, as they capture a non-trivial number of numerator events.

2b4. 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 2b5.

2b4.1. What method of controlling for differences in case mix is used?
☐ **No risk adjustment or stratification**
☒ **Statistical risk model with** 188 **risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.
Not applicable

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

Clinical Factors
For PSI 12, we considered a standard set of covariates grouped into four categories: demographics, severity of illness, comorbidities and transfer-in status. Covariates that were considered as potential risk adjusters included gender and age, MDC, Modified Diagnostic Related Groups (MDRGs) (defined as the base MS-DRG without comorbidity or complication distinctions), AHRQ Comorbidity Software categories and whether they were transferred from another facility. Only those covariates with at least 30 cases for PSI 12 are retained. A parsimonious model was identified using backward stepwise selection with bootstrapping.

The omitted covariate within mutually exclusive categories is the reference group for those categories. Reference categories are usually 1) the most common and/or 2) the least risk. The choice of omitted reference category does not affect predicted probabilities or model performance.

For the MDRGs, the risk reported is the residual risk after adjustment for the MDC to which the MDRG belongs. Likewise, the risk reported for MDCs represents the average risk of all MSDRGs in that MDC not included in the model.

Additional details are available in the *AHRQ Quality Indicator Empirical Methods* document, included in the supplemental file and available on the AHRQ QI website.

Sociodemographic Factors
Unlike other PSIs, racial differences in the incidence of PSI12 – indeed most manifestations of venous thromboembolism – have been recognized for at least 15 years. Specifically, Zakai & McClure (2011)[12] summarized in a recent review of the topic, "VTE appears to be most common in individuals of African descent in North America, with the incidence among Europeans in North America and Europe nearly as high, and a much lower incidence among people of Hispanic descent in the US, and Asian populations in both the US and Asia and Native Americans." Although the evidence regarding causal mechanisms is less clear, "genetics is thought to account for up to 60% of the risk of VTE, with the two known… single nucleotide polymorphisms associated with VTE risk (factor V Leiden and prothrombin gene polymorphism) found predominantly in European ancestry populations." However, variation in the prevalence of these polymorphisms is not sufficient to account for observed racial variation in incidence of VTE, leading these reviewers to conclude tentatively that "genetic predispositions to thrombosis, such as FV Leiden and prothrombin G20210A in European populations, and

[12] Zakai NA, McClure LA. Racial differences in venous thromboembolism. *J Thromb Haemost.* 2011;9(10):1877-1882.

high FVIII, high von Willebrand factor and low protein C in African populations, may play a role…". Accordingly, AHRQ has been actively evaluating this literature and considering its potential impact on PSI 12. Factors to consider are: (1) the quality and completeness of race/ethnicity reporting vary substantially across administrative data sets in the US; (2) extreme genetic diversity within the "black" and "Asian or Pacific Islander" categories that are used in US data sets; (3) concern that perceived racial variation in VTE incidence may become a "self-fulfilling prophecy" if it leads to variation in diagnostic practices (i.e., surveillance intensity); (4) most of the published data on this topic come from West Coast states and Hong Kong, where Han Chinese ancestry predominates; (5) racial differences appear to be lower in the setting of nearly universal pharmacologic prophylaxis (e.g., total hip or knee arthroplasty, valve replacement) than in other clinical settings[13]. Given the sensitivity of this topic, and the historical NQF-led consensus against race/ethnicity adjustment in publicly reported outcome measures, AHRQ is exploring the extent to which observed racial differences are consistent across states, and the extent to which adjusting for race would substantially change hospital-specific PSI 12 rates.

There is no evidence or causal model to suggest that socioeconomic factors other than race are associated with postoperative thromboembolic events independent of quality of care, or are mediated by pre-hospital care (which may not fall within the proper realm of hospital accountability). Accordingly, consistent with the guidance provided by NQF in the SDS Trial Period FAQs, AHRQ believes that it would be inappropriate to include other SDS variables in the risk-adjustment approach for PSI 12, which is an in-hospital outcome measure.

2b4.4a. What were the statistical results of the analyses used to select risk factors?

The current risk adjustment coefficients for PSI 12 can be found attached to the technical specifications document. The risk model includes 187 risk categories, including 26 age-gender categories in 5-year age categories between ages 30 and 89, and 2 age-gender categories ranging from below age 30 (i.e. 18-29) as one category and ages 90+ as another category, transfer in from another acute care facility and 13 comorbidities. The remainder of selected risk factors account for the reason for admission and the type of surgery that was performed during the hospitalization, including MDC and MS-DRGs collapsed to remove Complication or Comorbidity/ Major Complication or Comorbidity (CC/MCC) distinctions.

Table 6. Risk Adjustment Coefficients for PSI 12 Perioperative Pulmonary Embolism or Deep Vein Thrombosis Rate

Variable	Parameter	DF	Estimate	Standard Error	Wald Chi-Square	pr>Chi-Square
Intercept	Intercept	1	-5.2014	0.0856	3691.8563	<.0001
Sex | Age Demographics						
M_AgeCat_1	Male | Age 18 - 29	1	-0.1744	0.0861	4.1031	0.0428
M_AgeCat_2	Male | Age 30 - 34	1	0.0532	0.0949	0.3140	0.5752
M_AgeCat_3	Male | Age 35 - 39	1	-0.1109	0.0848	1.7124	0.1907
M_AgeCat_4	Male | Age 40 - 44	1	-0.1013	0.0678	2.2274	0.1356
M_AgeCat_5	Male | Age 45 - 49	1	-0.0281	0.0545	0.2654	0.6064
M_AgeCat_6	Male | Age 50 - 54	1	-0.0719	0.0453	2.5147	0.1128
M_AgeCat_7	Male | Age 55 - 59	1	-0.0331	0.0410	0.6550	0.4183
M_AgeCat_8	Male | Age 60 - 64	1	0.0334	0.0395	0.7133	0.3984
M_AgeCat_9	Male | Age 65 - 69	1	0.0582	0.0388	2.2504	0.1336
M_AgeCat_10	Male | Age 70 - 74	1	0.1541	0.0389	15.7288	<.0001
M_AgeCat_11	Male | Age 75 - 79	1	0.0117	0.0416	0.0794	0.7781
M_AgeCat_12	Male | Age 80 - 84	1	0.0401	0.0475	0.7132	0.3984
M_AgeCat_13	Male | Age 85 - 89	1	-0.0326	0.0605	0.2907	0.5898

[13] White RH, Keenan CR. Effects of race and ethnicity on the incidence of venous thromboembolism. *Thromb Res.* 2009;123 Suppl 4:S11-17.

Variable	Parameter	DF	Estimate	Standard Error	Wald Chi-Square	pr>Chi-Square
M_AgeCat_14	Male | Age >=90	1	-0.0181	0.0898	0.0406	0.8403
F_AgeCat_1	Female | Age 18 - 29	1	-0.1401	0.1039	1.8173	0.1776
F_AgeCat_2	Female | Age 30 - 34	1	-0.0052	0.1038	0.0025	0.9602
F_AgeCat_3	Female | Age 35 - 39	1	-0.0729	0.0880	0.6866	0.4073
F_AgeCat_4	Female | Age 40 - 44	1	-0.0433	0.0726	0.3564	0.5505
F_AgeCat_5	Female | Age 45 - 49	1	-0.1484	0.0604	6.0374	0.0140
F_AgeCat_6	Female | Age 50 - 54	1	-0.0779	0.0501	2.4180	0.1199
F_AgeCat_7	Female | Age 55 - 59	1	-0.1578	0.0440	12.8373	0.0003
F_AgeCat_8	Female | Age 60 - 64	1	-0.0515	0.0397	1.6799	0.1949
F_AgeCat_9	Female | Age 65 - 69	*Reference Group				
F_AgeCat_10	Female | Age 70 - 74	1	0.1101	0.0386	8.1115	0.0044
F_AgeCat_11	Female | Age 75 - 79	1	0.0848	0.0400	4.4993	0.0339
F_AgeCat_12	Female | Age 80 - 84	1	0.0687	0.0426	2.5999	0.1069
F_AgeCat_13	Female | Age 85 - 89	1	0.0227	0.0483	0.2209	0.6384
F_AgeCat_14	Female | Age >=90	1	-0.2748	0.0658	17.4196	<.0001
Origin						
TRNSFER	Transfer from another facility	1	0.5340	0.0260	420.4704	<.0001
Comorbidities						
ANEMDEF	Deficiency Anemias	1	0.0480	0.0204	5.5434	0.0186
BLDLOSS	Chronic blood loss anemia	1	0.2842	0.0581	23.9132	<.0001
CHF	Congestive heart failure	1	0.1131	0.0269	17.6364	<.0001
COAG	Coagulopathy	1	0.2406	0.0320	56.6425	<.0001
DEPRESS	Depression	1	-0.0508	0.0235	4.6875	0.0304
DM	Diabetes w/o chronic complications	1	-0.0763	0.0189	16.2695	<.0001
DMCX	Diabetes w/ chronic complications	1	-0.2402	0.0355	45.8000	<.0001
HTN_C	Hypertension	1	-0.0868	0.0159	29.8594	<.0001
HYPOTHY	Hypothyroidism	1	-0.0856	0.0221	15.0728	0.0001
IMMUNE	Immune disorders	1	0.4588	0.0259	312.7478	<.0001
LIVER	Liver disease	1	-0.2223	0.0457	23.6338	<.0001
LYMPH	Lymphoma	1	0.3274	0.0686	22.7664	<.0001
LYTES	Fluid and electrolyte disorders	1	0.2460	0.0200	151.6918	<.0001
METS	Metastatic cancer	1	0.7548	0.0283	709.2850	<.0001
OBESE	Obesity	1	0.3961	0.0186	455.4162	<.0001
PARA	Paralysis	1	0.4390	0.0390	126.8191	<.0001
PSYCH	Psychoses	1	0.2696	0.0364	54.7987	<.0001
PULMCIRC	Pulmonary circulation disease	1	0.7362	0.0363	410.8863	<.0001
RENLFAIL	Renal failure	1	-0.0923	0.0244	14.2972	0.0002
TUMOR	Solid tumor w/out metastasis	1	0.3911	0.0403	94.0014	<.0001
WGHTLOSS	Weight loss	1	0.3377	0.0264	164.1445	<.0001

Variable	Parameter	DF	Estimate	Standard Error	Wald Chi-Square	pr>Chi-Square
Major Diagnostic Categories (MDC)						
MDC_1	MDC 1: Nervous System	1	1.9541	0.1102	314.2587	<.0001
MDC_3	MDC 3: Ear Nose Mouth And Throat	1	-0.1993	0.1503	1.7582	0.1848
MDC_4	MDC 4: Respiratory System	1	1.4206	0.1015	195.9905	<.0001
MDC_5	MDC 5: Circulatory System	1	1.1995	0.1016	139.3844	<.0001
MDC_6	MDC 6: Digestive System	1	1.8675	0.1028	330.0538	<.0001
MDC_7	MDC 7: Hepatobiliary System And Pancreas	1	1.9318	0.1402	189.9777	<.0001
MDC_8	MDC 8: Musculoskeletal And Connective	1	1.8273	0.1125	263.7112	<.0001
MDC_9	MDC 9: Skin Subcutaneous And Breast	1	1.1874	0.2142	30.7385	<.0001
MDC_10	MDC 10: Endocrine Nutritional And Metabolic	1	1.5149	0.1543	96.3315	<.0001
MDC_11	MDC 11: Kidney And Urinary Tract	1	1.4906	0.1504	98.1772	<.0001
MDC_13	MDC 13: Female Reproductive System	1	1.8641	0.3079	36.6559	<.0001
MDC_16	MDC 16: Blood and Immunological	1	1.7823	0.1893	88.5996	<.0001
MDC_18	MDC 18: Infectious and Parasitic	1	1.0639	0.1130	88.5790	<.0001
MDC_20	MDC 20: Alcohol/Drug Disorders	1	2.0751	0.3936	27.7909	<.0001
MDC_21	MDC 21: Injuries Poison And Toxic	1	1.7901	0.1588	127.0329	<.0001
MDC_22	MDC 22: Burns	1	1.8130	0.2395	57.3105	<.0001
MDC_23	MDC 23: Factors Influencing Health	1	2.3070	0.7610	9.1904	0.0024
Modified Diagnostic Related Groups (MDRG)						
mdrg_1001	Adrenal & pituitary procedures	1	-1.7464	0.2079	70.5367	<.0001
mdrg_1002	Amputat of lower limb for endocrine	1	-2.9994	0.2193	187.0073	<.0001
mdrg_1003	O.R. procedures for obesity	1	-3.1502	0.1742	326.9698	<.0001
mdrg_1004	Skin grafts & wound debrid for endoc	1	-2.7900	0.2832	97.0474	<.0001
mdrg_1005	Thyroid parathyroid & thyroglossal procedures	1	-3.6907	0.2599	201.7233	<.0001
mdrg_1006	Other endocrine nutrit & metab proc	1	-2.1992	0.1883	136.3546	<.0001
mdrg_102	Cranio w major dev impl/acute complex CNS	1	-1.4508	0.1194	147.5879	<.0001
mdrg_103	Craniotomy & endovascular intracranial procedures	1	-2.0319	0.0920	487.8010	<.0001
mdrg_104	Spinal procedures	1	-2.1266	0.1224	302.0674	<.0001
mdrg_105	Ventricular shunt	1	-2.5795	0.1813	202.5275	<.0001

Variable	Parameter	DF	Estimate	Standard Error	Wald Chi-Square	pr>Chi-Square
	procedures					
mdrg_106	Carotid artery stent procedure	1	-4.3568	0.4155	109.9654	<.0001
mdrg_107	Extracranial procedures	1	-4.3151	0.1757	603.2585	<.0001
mdrg_108	Periph & cranial nerve & other nerv syst proc	1	-2.7697	0.1548	320.2558	<.0001
mdrg_1101	Kidney transplant	1	-2.5456	0.2148	140.4227	<.0001
mdrg_1102	Major bladder procedures	1	-1.0869	0.1534	50.1998	<.0001
mdrg_1103	Kidney & ureter procedures for neoplasm	1	-1.6111	0.1460	121.8085	<.0001
mdrg_1104	Kidney & ureter procedures for non-neoplasm	1	-2.1302	0.1518	196.8980	<.0001
mdrg_1105	Minor bladder procedures	1	-2.2142	0.2580	73.6545	<.0001
mdrg_1106	Prostatectomy	1	-2.7549	0.3417	65.0078	<.0001
mdrg_1107	Transurethral procedures	1	-2.8046	0.1656	286.6747	<.0001
mdrg_1108	Urethral procedures	1	-2.5711	0.4657	30.4820	<.0001
mdrg_1109	Other kidney & urinary tract procedures	1	-2.3278	0.1589	214.6566	<.0001
mdrg_1201	Major male pelvic procedures	1	-1.3894	0.1445	92.4233	<.0001
mdrg_1202	Penis procedures	1	-0.9246	0.3631	6.4832	0.0109
mdrg_1203	Testes procedures	1	-1.1998	0.4168	8.2857	0.0040
mdrg_1204	Transurethral prostatectomy	1	-1.4496	0.1972	54.0361	<.0001
mdrg_1301	Pelvic evisceration - rad hysterectomy	1	-1.7228	0.3226	28.5224	<.0001
mdrg_1302	Uterine & adnexa proc ovarian or adnexal malig	1	-0.9227	0.3076	8.9988	0.0027
mdrg_1303	Uterine adnexa proc non-ovarian/adnexal malig	1	-1.7884	0.3106	33.1530	<.0001
mdrg_1304	Uterine & adnexa proc for non-malignancy	1	-3.2187	0.3054	111.0827	<.0001
mdrg_1305	DnC conization laparoscopy & tubal interruption	1	-2.4327	0.3491	48.5627	<.0001
mdrg_1306	Vagina cervix & vulva procedures	1	-3.2510	0.3885	70.0195	<.0001
mdrg_1307	Female reproductive system reconstructive	1	-4.4391	0.5371	68.2985	<.0001
mdrg_1308	Other female reproductive system procedures	1	-1.5271	0.3336	20.9534	<.0001
mdrg_1601	Splenectomy	1	-1.3589	0.2369	32.8960	<.0001
mdrg_1602	Other O.R. proc of the blood & blood forming	1	-2.2902	0.2870	63.6580	<.0001
mdrg_1707	Lymphoma & leukemia	1	-0.3342	0.1597	4.3781	0.0364
mdrg_1708	Lymphoma & non-acute leukemia	1	0.0641	0.1310	0.2395	0.6246
mdrg_1801	Infectious & parasitic diseases w procedure	1	-1.0659	0.0847	158.2643	<.0001

Variable	Parameter	DF	Estimate	Standard Error	Wald Chi-Square	pr>Chi-Square
mdrg_1802	Postoperative or post-traumatic infections	1	-1.3481	0.1085	154.4261	<.0001
mdrg_2101	Wound debridements for injuries	1	-2.1844	0.2264	93.0519	<.0001
mdrg_2102	Skin grafts for injuries	1	-2.0398	0.2242	82.7521	<.0001
mdrg_2103	Hand procedures for injuries	1	-3.6041	0.5188	48.2556	<.0001
mdrg_2104	Other O.R. procedures for injuries	1	-2.1380	0.1521	197.5827	<.0001
mdrg_2201	Full thickness burn w skin graft or inhal inj	1	-2.1319	0.3018	49.8875	<.0001
mdrg_2210	Extensive burns or full thickness burns	1	-0.7971	0.3454	5.3265	0.0210
mdrg_2301	O.R. proc w diagnoses of other contact	1	-2.9807	0.7732	14.8629	0.0001
mdrg_2407	Limb reattachment hip & femur proc	1	1.4224	0.1057	181.0387	<.0001
mdrg_2408	Other O.R. procedures for multiple sig tr	1	1.2703	0.0998	161.8979	<.0001
mdrg_301	Acute major eye infections	1	-1.7501	0.2740	40.8001	<.0001
mdrg_302	Other ear nose mouth & throat O.R. procedures	1	-1.7744	0.3052	33.7988	<.0001
mdrg_304	Mouth procedures	1	-1.5320	0.4279	12.8197	0.0003
mdrg_305	Salivary gland procedures	1	-1.5616	0.7188	4.7191	0.0298
mdrg_401	Major chest procedures	1	-1.9713	0.0836	555.6369	<.0001
mdrg_402	Other resp system O.R. procedures	1	-1.9591	0.0890	484.3703	<.0001
mdrg_502	Perc cardiovasc proc w non-drug-eluting stent	1	-2.7420	0.1239	490.0206	<.0001
mdrg_503	Cardiac valve & oth maj cardiothoracic proc	1	-1.6849	0.0791	453.7381	<.0001
mdrg_504	Cardiac defibrillator implant	1	-2.4273	0.1224	393.5079	<.0001
mdrg_505	Other cardiothoracic procedures	1	-1.6255	0.1785	82.9056	<.0001
mdrg_506	Coronary bypass w PTCA	1	-1.2387	0.1938	40.8648	<.0001
mdrg_507	Coronary bypass w cardiac cath	1	-1.7959	0.0797	507.1579	<.0001
mdrg_509	Amputation for circ sys disorders	1	-1.7187	0.1123	234.1448	<.0001
mdrg_510	Permanent cardiac pacemaker implant	1	-2.4365	0.1652	217.4377	<.0001
mdrg_511	Perc cardiovasc proc w drug-eluting stent	1	-3.2142	0.0908	1252.7636	<.0001
mdrg_513	Perc cardiovasc proc w/o coronary artery stent	1	-2.6817	0.1213	488.3835	<.0001
mdrg_514	Other vascular procedures	1	-2.1306	0.0852	625.6172	<.0001
mdrg_515	Upper limb & toe amputation	1	-2.4288	0.2510	93.6017	<.0001
mdrg_516	Cardiac pacemaker device	1	-3.2909	0.5812	32.0652	<.0001

Variable	Parameter	DF	Estimate	Standard Error	Wald Chi-Square	pr>Chi-Square
	replacement					
mdrg_517	Cardiac pacemaker revision	1	-1.9891	0.1964	102.5808	<.0001
mdrg_519	Other circulatory system O.R. procedures	1	-2.5212	0.1459	298.7615	<.0001
mdrg_601	Stomach esophageal & duodenal	1	-1.8190	0.0782	540.8606	<.0001
mdrg_602	Major small & large bowel proce	1	-1.8828	0.0686	753.0419	<.0001
mdrg_603	Rectal resection	1	-2.4953	0.1371	331.3630	<.0001
mdrg_604	Peritoneal adhesiolysis	1	-2.0364	0.0852	571.9076	<.0001
mdrg_605	Appendectomy w complicated principal diag	1	-3.1089	0.1546	404.2923	<.0001
mdrg_606	Appendectomy w/o complicated principal diag	1	-4.4727	0.2071	466.3873	<.0001
mdrg_607	Minor small & large bowel procedures	1	-2.7722	0.1613	295.4891	<.0001
mdrg_608	Anal & stomal procedures	1	-3.4385	0.1884	332.9490	<.0001
mdrg_609	Inguinal & femoral hernia procedures	1	-2.8022	0.1585	312.4779	<.0001
mdrg_610	Hernia procedures except inguinal & femoral	1	-2.4488	0.0993	607.5484	<.0001
mdrg_611	Other digestive system O.R. procedures	1	-2.0970	0.1006	434.8410	<.0001
mdrg_701	Pancreas liver & shunt procedures	1	-1.8510	0.1328	194.1806	<.0001
mdrg_702	Biliary tract proc except only cholecyst	1	-2.0498	0.2166	89.5895	<.0001
mdrg_703	Cholecystectomy w c.d.e.	1	-2.9038	0.4248	46.7198	<.0001
mdrg_704	Cholecystectomy except by laparoscope	1	-2.6571	0.1668	253.7459	<.0001
mdrg_705	Laparoscopic cholecystectomy	1	-3.8349	0.1377	775.6323	<.0001
mdrg_706	Hepatobiliary diagnostic procedures	1	-2.4405	0.2526	93.3692	<.0001
mdrg_707	Other hepatobiliary or pancreas procedures	1	-2.2476	0.2284	96.8388	<.0001
mdrg_7701	Heart transplant or implant heart assist sys	1	-0.4843	0.1400	11.9676	0.0005
mdrg_7702	Liver transplant	1	-2.0613	0.2034	102.7390	<.0001
mdrg_7703	Lung transplant	1	-0.6787	0.1991	11.6179	0.0007
mdrg_801	Combined anterior/posterior spinal fusion	1	-1.6308	0.1107	216.9013	<.0001
mdrg_802	Spinal fus exc cerv w spinal curv/malig/infec	1	-1.1376	0.1098	107.3083	<.0001
mdrg_803	Spinal fusion except cervical	1	-2.4295	0.0911	711.1304	<.0001
mdrg_804	Bilateral or multiple major joint procs	1	-1.5441	0.1061	211.9112	<.0001

Variable	Parameter	DF	Estimate	Standard Error	Wald Chi-Square	pr>Chi-Square
mdrg_805	Wnd debrid & skn grft exc hand for musculo	1	-2.0722	0.1107	350.4738	<.0001
mdrg_806	Revision of hip or knee replacement	1	-2.2587	0.0985	525.9427	<.0001
mdrg_807	Major joint replacement or reattachment	1	-2.4109	0.0817	871.3145	<.0001
mdrg_808	Cervical spinal fusion	1	-3.4341	0.1189	834.2653	<.0001
mdrg_809	Amputation for musculoskeletal sys	1	-2.3886	0.1639	212.3057	<.0001
mdrg_810	Biopsies of musculoskeletal system	1	-2.2951	0.1262	330.7007	<.0001
mdrg_811	Hip & femur procedures except major joint	1	-1.9509	0.0850	526.4460	<.0001
mdrg_812	Major joint & limb reattachment	1	-3.2298	0.1312	605.8434	<.0001
mdrg_813	Knee procedures w pdx of infection	1	-2.5749	0.1940	176.1050	<.0001
mdrg_814	Knee procedures w/o pdx of infection	1	-2.6819	0.1705	247.2878	<.0001
mdrg_815	Back & neck proc exc spinal fusion	1	-3.0030	0.1064	796.0842	<.0001
mdrg_816	Lower extrem & humer proc	1	-2.6531	0.0975	740.8094	<.0001
mdrg_817	Local excision & removal int fix devices	1	-3.1469	0.1807	303.3972	<.0001
mdrg_818	Local excision & removal int fix devices	1	-2.2297	0.2707	67.8452	<.0001
mdrg_819	Soft tissue procedures	1	-2.7400	0.1428	368.1118	<.0001
mdrg_820	Foot procedures	1	-3.1483	0.2277	191.2020	<.0001
mdrg_821	Major thumb or joint	1	-3.5616	0.7119	25.0265	<.0001
mdrg_822	Major shoulder or elbow joint procedures	1	-2.6603	0.3867	47.3191	<.0001
mdrg_824	Shoulder elbow or forearm proc	1	-3.2033	0.1939	272.8594	<.0001
mdrg_825	Hand or wrist proc	1	-4.0345	0.4161	94.0180	<.0001
mdrg_826	Other musculoskelet sys & conn tiss proc	1	-2.4057	0.1205	398.7629	<.0001
mdrg_8899	Non-Extensive O.R. Proc Unrelated to PDX	1	-2.2699	0.0604	1411.4042	<.0001
mdrg_901	Skin graft &/or debrid for skn ulcer or cellulitis	1	-2.4763	0.2382	108.0504	<.0001
mdrg_902	Skin graft &/or debrid exc for skin ulcer	1	-2.1898	0.2884	57.6372	<.0001
mdrg_903	Other skin subcut tiss & breast	1	-2.9103	0.2334	155.5140	<.0001
mdrg_904	Mastectomy for malignancy	1	-3.1344	0.3262	92.3472	<.0001
mdrg_905	Breast biopsy local excision	1	-2.8499	0.3885	53.8104	<.0001

c-statistic = .751

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

Not applicable (see above)

2b4.5. 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*)

This analysis evaluates how strongly the risk adjustment model is associated with the event of interest. The measure of discrimination, how well the risk adjustment model distinguishes events from non-events, is the c-statistic (also known as the area under a receiver operating characteristic curve). The c-statistic is computed by assigning each observation a predicted probability of the outcome from the risk-adjustment model, based on the value of the observed covariates and the parameter estimates from the risk-adjustment model. Two copies of the dataset are sorted, first from highest to lowest predicted probability and second from lowest to highest predicted probability. Random sampling is used to create a set of paired observations. Pairs that consist of one event and one non-event (discordant pairs) are kept and concordant pairs are discarded. The c-statistic is a measure of the proportion of discordant pairs of observations for which the observation with the event had a higher predicted probability from the risk-adjustment model than the observation without the event. C-statistics above 0.70 and below 0.80 have moderate discrimination. Above 0.80 the discrimination is high. We did not employ common "goodness of fit" tests because these tests tend to be uninformative with large samples.

We also evaluated the calibration of the risk adjustment model by evaluating how closely observed and predicted rates compare across deciles of the predicted rate. This analysis splits the sample into deciles based on predicted rates, and then compares these rates with the observed rates for the population in each decile. A well calibrated model, or one that does not over or under-estimate risk, will have comparable observed and predicted rates across the risk spectrum.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

Table 7. Risk adjustment Model Discrimination and Calibration PSI 12 Perioperative Pulmonary Embolism and Deep Vein Thrombosis Rate in in 2-year Pooled Data (2012-2013)

Predicted Rate Decile	Number of Discharges per Decile	Predicted Rate (per 1,000 surgical discharges)	Observed Rate (per 1,000 surgical discharges)	Observed to Predicted Ratio
1 (lowest)	1,138,612	0.5156	0.3794	0.736
2	1,138,613	0.8409	0.7281	0.866
3	1,138,613	1.2150	1.1584	0.953
4	1,138,613	1.7430	1.7556	1.007
5	1,138,613	2.3928	2.3432	0.979
6	1,138,613	2.9339	2.9554	1.007
7	1,138,613	3.4817	3.5447	1.018
8	1,138,613	4.2860	4.4783	1.045
9	1,138,613	5.7039	6.2655	1.098
10 (highest)	1,138,613	13.3531	14.4211	1.080
C-Statistic	0.751			

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2012 - 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov (AHRQ QI Software Version 6.0)

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

See Table 7 in 2b4.6

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:
See calibration by decile in Table 7 in 2b4.6

2b4.9. Results of Risk Stratification Analysis:

Not applicable

2b4.10. 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 risk-adjustment model has moderately high discrimination, based on a c statistic of 0.751 (i.e., in 75% of randomly selected pairs of discordant observations, the patient who experienced PSI 12 had a higher probability of experiencing the event than the patient who did not). A model that is well calibrated will have observed values similar to predicted values across the predicted value deciles. This indicator is well calibrated, as the observed to predicted ratio values across the deciles range between 0.87 to 1.09 for all deciles except the lowest decile. For patients with very low predicted rates, the relative difference between observed and predicted values is greater, but not particularly concerning due to the very small number of events that occur in this risk stratum.

2b4.11. 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*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. 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)*

This analysis assesses the probability that a hospital is higher or lower than a benchmark or threshold, given hospital size. It reflects whether the indicator can discriminate the best performing hospitals from the lower performing hospitals.

For this analysis, "benchmark" refers to the smoothed indicator rate based on the 20th percentile of the reference population (i.e., 20% of hospitals have a lower mortality rate or better performance). "Threshold" refers to the indicator rate based on the 80th percentile (i.e., 80% have lower mortality or better performance). Assuming an underlying Gamma distribution for the smoothed rates of the measure, the benchmark and threshold values are identified using population reference rates and signal variances computed from the entire AHRQ QI POA Reference Population. Hospital-level 90% confidence limits for smoothed rates are also computed from the Gamma distribution.

The analysis is reported by size decile, based on the denominator cases, demonstrating performance across hospitals of various sizes. Each hospital is assumed to have an underlying distribution of smoothed rates that follows a Gamma distribution. The parameters of a Gamma distribution are shape and scale. For each hospital the shape is calculated as ((smoothed rate)^2/ smoothed rate variance), and the scale is calculated as (smoothed rate variance / smoothed rate). The smoothed rate variance (aka posterior variance) is calculated as the signal variance – (reliability weight * signal variance). The reliability weight is calculated as (signal variance / (signal variance + noise variance)). Hospitals are ranked by size and grouped into 10 equal categories of size (deciles). The Benchmark and Threshold are compared to the Gamma distribution of the smoothed rates for each hospital to determine if the hospital rate is better or worse than the Benchmark and Threshold rates with 95% probability. This provides a 95% confidence interval for the Benchmark and Threshold rate.

Table 8 reports the proportion of hospitals above (better than) and below (worse than) the Benchmark and Threshold rates and the proportion not classified as either above or below. The proportion of hospitals not classified as either better or worse have rates that fall within the 95% confidence interval.

2b5.2. 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. Performance Categories by Hospital Size Decile PSI 12 Postoperative Pulmonary Embolism and Deep Vein Thrombosis Rate Using in 2-year Pooled Data (2012-2013)

			Benchmark			Threshold		
Size Decile	Number of Hospitals	Average Number of Denominator Discharges Per Hospital	Proportion Better	Proportion Worse	Proportion Unclassified	Proportion Better	Proportion Worse	Proportion Unclassified
1 (smallest)	357	45.3	0.0000	0.0112	0.9888	0.0000	0.0000	1.0000
2	358	228.5	0.0000	0.0754	0.9246	0.0000	0.0000	1.0000
3	357	541.9	0.0000	0.1457	0.8543	0.0056	0.0000	0.9944
4	358	998.3	0.0000	0.1899	0.8101	0.2235	0.0000	0.7765
5	357	1,608.1	0.0000	0.2465	0.7535	0.3725	0.0140	0.6134
6	358	2,372.7	0.0000	0.3436	0.6564	0.4553	0.0196	0.5251
7	358	3,302.9	0.0000	0.4190	0.5810	0.4888	0.0335	0.4777
8	357	4,548.3	0.0000	0.5210	0.4790	0.5210	0.0532	0.4258
9	358	6,561.0	0.0056	0.5642	0.4302	0.5726	0.0754	0.3520
10 (largest)	357	11,648.9	0.0028	0.7759	0.2213	0.5266	0.1289	0.3445
Overall	3,575	3,184.9	0.0008	0.3292	0.6699	0.3166	0.0324	0.6509

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2012 - 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov (AHRQ QI Software Version 6.0)

2b5.3. 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?*)

Over all hospitals, this indicator has modest discrimination for identifying low or high performing hospitals; 35% of hospitals can be classified as better or worse than the threshold (the percentage classified as either above or below the threshold) and 33% better or worse than the benchmark (the percentage classified as either above or below the benchmark). However, as hospital size increases, the discrimination also increases such that for hospitals in the largest 3 deciles the algorithm classifies 57% - 66% of hospitals against the threshold and 52%-78% of hospitals against the benchmark.

2b6. 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 SDS 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 SDS 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.

Not applicable

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)
Not applicable

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)
Not applicable

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)
Not applicable

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

The AHRQ QIs use frequently reported administrative data variables. PSI 12 excludes cases with missing discharge disposition, age, sex, discharge quarter, discharge year, and principal diagnosis. These variables are required for indicator construction and are required of all hospital discharge records. The frequency of missing data for each variable is available by state and year from the AHRQ HCUP website (http://www.hcup-us.ahrq.gov/cdstats/cdstats_search.jsp).

2b7.2. 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*)

For these variables, frequencies of missing data are typically less than 1% of the state database. It is unlikely the bias would occur from such a low frequency of missing data.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

Exclusion of cases with missing data for these variables is appropriate.

3. Feasibility

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

3a. Byproduct of Care Processes

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

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)
If other:

3b. Electronic Sources

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

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*
ALL data elements are in defined fields in electronic claims

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

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.**
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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
Because the indicator is based on readily available administrative billing and claims data, feasibility is not an issue. This version of the indicator requires present-on-admission (POA) data for risk-adjustment and for specification of the numerator and denominator. POA indicators were added as data elements to the uniform bill form (UB-04) effective October 1, 2007. Hospitals incurred a payment penalty for not including POA status on Medicare records beginning October 1, 2008. Each of the secondary diagnoses in a discharge record can be flagged as "present at the time the order for inpatient admission occurs" or not (see http://www.cdc.gov/nchs/icd/icd9cm_addenda_guidelines.htm). The number of states reporting consistent POA has increased dramatically since 2008.

The AHRQ QI software has been publicly available at no cost since 2001; Users have over ten years of experience using the AHRQ QI software in SAS and Windows.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
There are no fees. Software is freely available from the AHRQ Quality Indicators website (http://www.qualityindicators.ahrq.gov/).

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.

Planned	Current Use (for current use provide URL)
	Public Reporting Arizona Department of Health Services, AZ Hospital Compare, MONAHRQ website http://pub.azdhs.gov/hospital-discharge-stats/2012/AboutQualityRatings.html#J CareChex (Division of Quantros) http://www.carechex.com/QualityIndicators.aspx Cigna Centers of Excellence Hospital Value Tool http://www.cigna.com/pdf/CentersOfExcellence.pdf CMS Medicare Hospital Compare Program https://www.medicare.gov/hospitalcompare/Data/Measures-Displayed.html# Colorado Hospital Association http://www.cohospitalquality.org/corda/dashboards/COLORADO_REPORT_CARD_BY_MEASURE/main.dashxml Commonwealth Fund, Why Not the Best http://whynotthebest.org/methodology Connecticut Department of Health Services, CT Hospital Compare, MONAHRQ website http://ctmonahrq.ct.gov/2012/index.html#/resources/AboutQualityRatings Connecticut Hospital Association http://www.cthosp.org/advocacy/quality-and-patient-safety/hospital-quality-reporting-website/ Florida Agency for Health Care Administration www.floridahealthfinder.gov Healthcare Association of New York State http://www.hanys.org/quality/data/report_cards/2013/docs/2013_hanys_report_card_book.pdf HealthGrades https://d2dcgio3q2u5fb.cloudfront.net/54/98/f79cdfd84640a03792ea092f20a8/2014-patient-safety-methodology.pdf Hospital Safety Score http://www.hospitalsafetyscore.org/media/file/HospitalSafetyScore_ScoringMethodology_Spring2015_Final.pdf Illinois Department of Public Health http://healthcarereportcard.illinois.gov/methodology Iowa Healthcare Collaborative https://iowareport.ihconline.org/Public/Reports.aspx?FID=778&F1ID=0&F2ID=0&F3ID=0&CID=2&PID=4 Kentucky Cabinet for Health and Family services https://prd.chfs.ky.gov/MONAHRQ/2012/MONAHRQ/AboutQualityRatings.html Kentucky Hospital Association Quality Data http://info.kyha.com/QualityData/ Louisiana Hospital Inform http://lahospitalinform.org/index.html Maine Health Data Organization (MHDO), MONAHRQ Website https://mhdo.maine.gov/monahrq/#/resources/AboutQualityRatings Maryland Health Care Commission, MONAHRQ Website http://www.hscrc.state.md.us/documents/md-maphs/wg-meet/di/2014-03-04/MHCC%20Inpatient%20Measures%20Inventory%20QBR%20highlights.pdf Minnesota Community Measurement http://mncm.org/reports-and-websites/reports-and-data/#-available-data Nevada Compare Care, MONAHRQ website http://nevadacomparecare.net/MQ2014/index.html#/professional/resources/AboutQualityRatings

	Nevada Hospital Association http://www.nvhospitalquality.net/old-home New Jersey Department of Health http://web.doh.state.nj.us/apps2/hpr/docs/2012/technicalreport_psi.pdf Niagara Health Quality Coalition, New York State Hospital Report Card http://www.myhealthfinder.com/newyork15/main_byproc.php Norton Healthcare http://www.nortonhealthcare.com/QualityReport Oklahoma State Department of Health, MONAHRQ https://www.phin.state.ok.us/ahrq/MONAHRQ%202010/Methodology.html South Dakota Association of Healthcare Organizations http://www.sdhospitalquality.org/search.php Texas Health Resources https://www.texashealth.org/Documents/System/Quality_Patient_Safety/Reports/03-02-2016_Surgery.pdf Think About It Colorado http://www.cohospitalquality.org/corda/dashboards/COLORADO_REPORT_CARD_BY_HOSPITAL/main.dashxml#cordaDash=1005 U.S. News and World Report http://www.usnews.com/pubfiles/BH2015-16MethodologyReport.pdf Utah Department of Health, MONAHRQ website https://health.utah.gov/myhealthcare/monahrq/ Virginia Health Information http://www.vhi.org/MONAHRQ/default.asp?yr=2013 Washington State, MONAHRQ website http://www.wamonahrq.net/MONAHRQ_5p0_WA_2012/index.html#/resources/AboutQualityRatings WHA Information Center (Wisconsin Hospital Association) http://www.whainfocenter.com/uploads/PDFs/Publications/QualityIndicators/2012_WI_IQIReport.pdf Quality Improvement with Benchmarking (external benchmarking to multiple organizations) CMS Hospital Compare http://www.medicare.gov/hospitalcompare/Data/Measures-Displayed.html University HealthSystem Consortium/Vizient https://www.vizientinc.com/clinical-analytics-and-benchmarking.htm Quality Improvement (Internal to the specific organization) Greenville Health System, Quality and Safety Report http://www.ghs.org/upload/docs/Reports/2013-April-Quality-Report.pdf Northwestern Memorial Hospital, Patient Safety Indicator Monitoring Plan http://www.nmh.org/nm/quality-bleeding-or-bruising-following-surgery Upstate University Hospital http://qoc.upstate.edu/QualityOfCare.cfm?quality_measure_group_id=7

4a.1. **For each CURRENT use, checked above, provide:**

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

Public Reporting
Arizona Department of Health Services, AZ Hospital Compare, MONAHRQ website
Hospital quality ratings from all hospitals in Arizona
http://pub.azdhs.gov/hospital-discharge-stats/2012/AboutQualityRatings.html#J

CareChex (Division of Quantros)
Provides comprehensive reports of hospitals to consumers, providers and purchasers

http://www.carechex.com/QualityIndicators.aspx

Cigna
Centers of Excellence Hospital Value Tool – Health insurance company
http://www.cigna.com/pdf/CentersOfExcellence.pdf

CMS Medicare Hospital Compare Program
Publically available database containing information about the quality of care at over 4,000 Medicare-certified hospitals across the U.S.
https://www.medicare.gov/hospitalcompare/Data/Measures-Displayed.html#

Colorado Hospital Association
Hospital quality ratings from hospitals in Colorado
http://www.cohospitalquality.org/corda/dashboards/COLORADO_REPORT_CARD_BY_MEASURE/main.dashxml

Commonwealth Fund, Why Not the Best
Provides performance and quality ratings for most US hospitals
http://whynotthebest.org/methodology

Connecticut Department of Health Services, CT Hospital Compare, MONAHRQ website
Hospital quality ratings from all hospitals in Connecticut.
http://ctmonahrq.ct.gov/2012/index.html#/resources/AboutQualityRatings

Connecticut Hospital Association
Provide quality of care for hospitals in Connecticut
http://www.cthosp.org/advocacy/quality-and-patient-safety/hospital-quality-reporting-website/

Florida Agency for Health Care Administration
Provide quality of care ratings from hospitals within Florida
www.floridahealthfinder.gov

Healthcare Association of New York State
Supports availability of hospital quality and safety information to help patients make choices and assist providers in improving care
http://www.hanys.org/quality/data/report_cards/2013/docs/2013_hanys_report_card_book.pdf

HealthGrades
Healthgrades measures 40 million patient records from 4,500 hospitals nationwide for the most recent three-year period. Consumer-targeted hospital and provider ratings
https://d2dcgio3q2u5fb.cloudfront.net/54/98/f79cdfd84640a03792ea092f20a8/2014-patient-safety-methodology.pdf

Hospital Safety Score
PSI 12 is one component of a single composite score that represents a hospital's overall performance in patient safety
http://www.hospitalsafetyscore.org/media/file/HospitalSafetyScore_ScoringMethodology_Spring2015_Final.pdf

Illinois Department of Public Health
Provides access to information on hospital and safety data in hospitals in Illinois
http://healthcarereportcard.illinois.gov/methodology

Iowa Healthcare Collaborative
Hospital quality ratings from hospitals in Iowa
https://iowareport.ihconline.org/Public/Reports.aspx?FID=778&F1ID=0&F2ID=0&F3ID=0&CID=2&PID=4

Kentucky Cabinet for Health and Family services
Hospital quality ratings from hospitals in Kentucky
https://prd.chfs.ky.gov/MONAHRQ/2012/MONAHRQ/AboutQualityRatings.html

Kentucky Hospital Association Quality Data

Hospital quality ratings from most hospitals in Kentucky
http://info.kyha.com/QualityData/

Louisiana Hospital Inform
Hospital quality ratings from hospitals in Louisiana
http://lahospitalinform.org/index.html

Maine Health Data Organization (MHDO), MONAHRQ Website
Hospital quality ratings from all hospitals in Maine
https://mhdo.maine.gov/monahrq/#/resources/AboutQualityRatings

Maryland Health Care Commission, MONAHRQ Website
Collects and provides quality ratings on hospitals across Maryland
http://www.hscrc.state.md.us/documents/md-maphs/wg-meet/di/2014-03-04/MHCC%20Inpatient%20Measures%20Inventory%20QBR%20highlights.pdf

Minnesota Community Measurement
Minnesota Community Measurement is a nonprofit healthcare data reporting organization. Provides quality ratings on hospitals across Minnesota.
http://mncm.org/reports-and-websites/reports-and-data/#-available-data

Nevada Compare Care, MONAHRQ website
Hospital quality ratings from most hospitals in Nevada
http://nevadacomparecare.net/MQ2014/index.html#/professional/resources/AboutQualityRatings

Nevada Hospital Association
Transparency and Performance: Demonstrates Nevada hospitals activity relating to specific clinical indicators.
http://www.nvhospitalquality.net/old-home

New Jersey Department of Health
Public report of PSI performance for New Jersey Hospital
http://web.doh.state.nj.us/apps2/hpr/docs/2012/technicalreport_psi.pdf

Niagara Health Quality Coalition, New York State Hospital Report Card
Consumer focused public report of quality indicator performance for NY hospitals.
http://www.myhealthfinder.com/newyork15/main_byproc.php

Norton Healthcare
Report patient satisfaction scores in Norton Healthcare hospitals and their performance on nationally recognized quality indicators and practices http://www.nortonhealthcare.com/QualityReport

Oklahoma State Department of Health, MONAHRQ
Compares quality ratings on hospitals across Oklahoma
https://www.phin.state.ok.us/ahrq/MONAHRQ%202010/Methodology.html

South Dakota Association of Healthcare Organizations
Use PSI 12 in a composite of serious complications in report of Oregon hospital quality.
http://www.sdhospitalquality.org/search.php

Texas Health Resources
Provides quality and safety reports for all Texas Health Resources
https://www.texashealth.org/Documents/System/Quality_Patient_Safety/Reports/03-02-2016_Surgery.pdf

Think About It Colorado
Report hospital quality for all hospitals in Colorado
http://www.cohospitalquality.org/corda/dashboards/COLORADO_REPORT_CARD_BY_HOSPITAL/main.dashxml#cordaDash=1005

U.S. News and World Report
National publication that lists ratings of U.S. medical centers based on performance
http://www.usnews.com/pubfiles/BH2015-16MethodologyReport.pdf

Utah Department of Health, MONAHRQ website
Report hospital quality for all hospitals in Utah
https://health.utah.gov/myhealthcare/monahrq/

Virginia Health Information
Compares quality ratings on hospitals across Virginia
http://www.vhi.org/MONAHRQ/default.asp?yr=2013

Washington State, MONAHRQ website
Information system of inpatient care utilization, quality, and potentially avoidable stays in Washington State's community hospitals
http://www.wamonahrq.net/MONAHRQ_5p0_WA_2012/index.html#/resources/AboutQualityRatings

WHA Information Center (Wisconsin Hospital Association)
Wisconsin Inpatient Hospital Quality Indicators Report
http://www.whainfocenter.com/uploads/PDFs/Publications/QualityIndicators/2012_WI_IQIReport.pdf

Quality Improvement (Internal to the specific organization)
Greenville Health System, Quality and Safety Report
All data was collected from four hospitals in the Greenville Health system and compared with internal rates
http://www.ghs.org/upload/docs/Reports/2013-April-Quality-Report.pdf

Northwestern Memorial Hospital, Patient Safety Indicator Monitoring Plan
Quality improvement initiative at 894-bed academic hospital
http://www.nmh.org/nm/quality-bleeding-or-bruising-following-surgery

Upstate University Hospital
Report of hospital rates against national benchmark (published online)
http://qoc.upstate.edu/QualityOfCare.cfm?quality_measure_group_id=7

Quality Improvement (external benchmarking to multiple organizations)
CMS Hospital Compare
Publically available performance measures for hospitals
http://www.medicare.gov/hospitalcompare/Data/Measures-Displayed.html

University HealthSystem Consortium/Vizient
Internal quality improvement efforts, documentation, and evaluation of AHRQ PSIs for quality improvement by its members
https://www.vizientinc.com/clinical-analytics-and-benchmarking.htm

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

4a.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.*)
N/A

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

See Table 1 in response to question 1b.2 for 2-year pooled rates (also included in supplemental materials). Additional data discussed below.

Nationwide rates of this measure have decreased steadily during 2011-2013 from 4.0 to 3.7 cases/1,000 hospitalizations at risk. This decrease is consistent with the 43% decrease in hospital-associated VTE between 2010 and 2014 (i.e., an interval of 4 years versus 2 years) reported from the Medicare Patient Safety Monitoring System (MPSMS) (http://www.ahrq.gov/sites/default/files/publications/files/interimhacrate2014_2.pdf and Hackbarth et al., 2014). Given that MPSMS data are based on detailed medical record review by trained nurses, this decrease is likely to reflect true improvements in care, although diagnostic or documentation practices may also have changed during this interval.

4b.2. **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.**

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
There is no definite evidence of unintended consequences for this measure. However, several recent papers have focused on the problem of surveillance bias, or variation in the incidence of VTE across hospitals that may be attributable to screening and diagnostic practices. Chung et al. (2015) and Minami and Bilimoria (2015) summarized several studies demonstrating that hospital testing practices, and determinants of those practices, are associated with both PSI12 rates and postoperative VTE rates based on National Surgical Quality Improvement Program (NSQIP) data. Holcomb et al. (2015) studied 25,975 patients meeting the criteria for the Surgical Care Improvement Project (SCIP)-VTE measures at 79 VA facilities and reported a positive correlation between inpatient surveillance and inpatient VTE rates at the hospital level (R=0.33, P=.003) but no significant correlation of inpatient surveillance with either postdischarge surveillance (R=0.11, P=.29) or postdischarge VTE rates (R=0.03, P=.76). These studies suggest that variation in testing practices may contribute to variation in PSI12 rates across hospitals, but it remains unclear whether these data reflect underdiagnosis of VTE at low-testing hospitals, overdiagnosis at high-testing hospitals, or the true incidence of symptomatic VTE. Use of PSI12 may inappropriately reward undertesting, but it may also appropriately penalize overtesting and overdiagnosis. Discouraging overdiagnosis and overtreatment of clinically insignificant VTE would be a desirable consequence of using PSI 12, because treatment of VTE is associated with a significant risk of hemorrhagic complications. Excluding distal DVTs, as AHRQ did in its V6 modification of the PSI12 numerator specification, may resolve this concern.

Chung JW, Ju MH, Kinnier CV, Sohn MW, Bilimoria KY. Postoperative venous thromboembolism outcomes measure: analytic exploration of potential misclassification of hospital quality due to surveillance bias. Ann Surg. 2015;261(3):443-444.
Minami CA, Bilimoria KY. Are Higher Hospital Venous Thromboembolism Rates an Indicator of Better Quality?: Evaluation of the Validity of a Hospital Quality Measure. Adv Surg. 2015;49:185-204.
Holcomb CN, DeRussy A, Richman JS, Hawn MT. Association Between Inpatient Surveillance and Venous Thromboembolism Rates After Hospital Discharge. JAMA Surg. 2015;150(6):520-527.

Coding professionals follow detailed guidelines, are subject to training and credentialing requirements, peer review and audit.

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
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 completely harmonized?

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.)
Not applicable

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** PSI_12_Supplemental_File_160513-635987585688667447.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Agency for Healthcare Research and Quality
Co.2 **Point of Contact:** Pamela, Owens, Pam.Owens@ahrq.hhs.gov, 301-427-1412-
Co.3 **Measure Developer if different from Measure Steward:** Agency for Healthcare Research and Quality
Co.4 **Point of Contact:** Mamatha, Pancholi, Mamatha.Pancholi@ahrq.hhs.gov, 301-427-1470-

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.
In 2002, two workgroups were convened to provide feedback on key indicator development decisions and methodology, including the usefulness of Perioperative Pulmonary Embolism or Deep Vein Thrombosis Rate (PSI 12), formerly known as Postoperative Pulmonary Embolism or Deep Vein Thrombosis Rate (PSI 12). These workgroups included a multispecialty panel and a surgical specialty panel; the active members were:

Charles Bethea, MD, Cardiologist
Oklahoma City, OK
Duke Clinical Research Institute
Nominated by the American College of Cardiology

John Hunt, MD, MPH, Trauma surgeon, critical care
New Orleans, LA
Health Science Center - Louisiana State University
Nominated by the American College of Surgeons

Franco Laghi, MD, Critical care physician
Maywood, IL
Loyola University
Nominated by the American Thoracic Society

John Nelson, MD, FACP, Internist/Hospitalist
Bellevue, WA
Overlake Hospital Medical Center
Nominated by the National Association of Inpatient Physicians

Carol A. Petersen, RN, BSN, MAOM, CNOR, Perioperative nursing specialist
Denver, CO
Center for Nursing Practice
Nominated by the Association of Peri-Operative Registered Nurses

Bruce Williams, MSN, RN, Critical care nurse specialist
Orangeburg, SC
The Regional Medical Center - of Orangeburg and Calhoun Counties
Nominated by the American Association of Critical-Care Nurses

Preston Winters, MD, FACP, Internist
White Plains, NY
White Plains Hospital Center
Nominated by the American College of Physicians

Rodney Appell, MD, urologist
Houston, TX
Baylor College of Medicine
Nominated by the American Urologic Association

Alan Freeland, MD, Orthopedic surgeon
Jackson, MS
University of Mississippi Medical Center
Nominated by the American Academy of Hand Surgeon)

Patricia Howson, MD, MSc, Orthopedic surgeon
Redwood City, CA
Kaiser Permanente
Nominated by the American Academy of Orthopedic Surgeons

William Hozak, MD, Orthopedic surgeon
Philadelphia, PA
Jefferson Medical School
Nominated by the American Association of Hip and Knee Surgeons

Mathew Indeck, MD, General Surgeon -trauma surgery

Danville, PA
Jefferson College of Medicine
Nominated by the American College of Surgeons

Bruce Kaufman, MD, Pediatric neurosurgeon
Milwaukee, WI
Medical College of Wisconsin
Nominated by the American Association of Neurological Surgeons

In 2013, ten panels of experts were convened to support the process of converting the AHRQ QIs from ICD-9-CM to ICD-10-CM/PCS in an accurate and transparent manner, to improve the validity and usefulness of the QIs. One of these panels –focused on critical care conditions - advised AHRQ on the ICD-10-CM/PCS specifications for PSI 12. The active members of this panel were:

Bradley Chipps, MD
Sacramento, CA
Capital Allergy and Respiratory Disease Center

Brian A. Cason, MD
San Francisco, CA
Department of Anesthesia and Perioperative Care
University of California, San Francisco and Veterans Affairs Medical Center

Colleen Stalvey, RHIT
Los Angeles, CA
AHIMA Approved ICD-10-CM/PCS Trainer
Cedars-Sinai Medical Center HIM Department

Patricia Anania Firouzan, MSIS, RHIA
Pittsburgh, PA
AHIMA Approved ICD-10-CM/PCS Trainer
University of Pittsburgh
HIM Dept, School of Health & Rehab Sciences

Jeanine Baskin, RN, BSN, CPHQ
Winston-Salem, NC
Novant Health, Clinical Quality Performance

Theresa Smiley, RN, CPHQ
Winston-Salem, NC
Novant Health, Clinical Quality Performance

Vicky A. Mahn-DiNicola RN,MS,CPHQ
Tucson, AZ
Healthcare Provider Solutions Group
Midas+ Solutions, A Xerox Company

Sandra Strack Arabian, CSTR, CAISS, EMT
Boston, MA
Tufts Medical Center
Division of Trauma and Acute Care Surgery

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2002
Ad.3 **Month and Year of most recent revision:** 06, 2016
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?** 06, 2016

Ad.7 **Disclaimers:** None
Ad.8 **Additional Information/Comments:** None

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 2909
De.2. **Measure Title:** Perioperative Hemorrhage or Hematoma Rate (PSI 09)
Co.1.1. **Measure Steward:** Agency for Healthcare Research and Quality
De.3. **Brief Description of Measure:** Perioperative hemorrhage or hematoma cases involving a procedure to treat the hemorrhage or hematoma, following surgery per 1,000 surgical discharges for patients ages 18 years and older. Excludes cases with a diagnosis of coagulation disorder; cases with a principal diagnosis of perioperative hemorrhage or hematoma; cases with a secondary diagnosis of perioperative hemorrhage or hematoma present on admission; cases where the only operating room procedure is for treatment of perioperative hemorrhage or hematoma; obstetric cases.
1b.1. **Developer Rationale:** Clinically, this indicator is intended to capture preventable and significant perioperative hemorrhage or hematoma events that are in excess of what is expected for the surgery type. The intent is to capture the significant events, such as those that are severe or where there may be a delay in diagnosis or treatment requiring reoperation, as these events are associated with a significant increase in harm to the patient.

S.4. **Numerator Statement:** Discharges, among cases meeting the inclusion and exclusion rules for the denominator, with:
• any secondary ICD-9-CM or ICD-10-CM diagnosis codes for perioperative hemorrhage or hematoma and any-listed ICD-9-CM or ICD-10-PCS procedure codes for treatment of hemorrhage or hematoma

Note that the ICD-10-CM specification is limited to postoperative hemorrhage or hematoma, whereas the ICD-9-CM specification captures both intraoperative and postoperative hemorrhage or hematoma (due to diagnosis codes that are less specific).
S.7. **Denominator Statement:** Surgical discharges, for patients ages 18 years and older, with any-listed ICD-9-CM or ICD-10-PCS procedure codes for an operating room procedure. Surgical discharges are defined by specific MS-DRG codes.

See Appendices: (attached in S.2b)
- Appendix A – Operating Room Procedure Codes
- Appendix E – Surgical Discharge MS-DRGs (for discharges on or after October 1, 2007)

S.10. **Denominator Exclusions:** Exclude cases:
• with a principal ICD-9-CM or ICD-10-CM diagnosis code (or secondary diagnosis present on admission(1) for perioperative hemorrhage or postoperative hematoma
• where the only operating room procedure is for treatment of perioperative hemorrhage or hematoma
• with any secondary ICD-9-CM or ICD-10-CM diagnosis codes for perioperative hemorrhage or hematoma and any-listed ICD-9-CM or ICD-10-PCS procedure codes for treatment of perioperative hemorrhage or hematoma occurring before the first operating room procedure(2)
• with any-listed ICD-9-CM or ICD-10-CM diagnosis codes for coagulation disorder
• MDC 14 (pregnancy, childbirth, and puerperium)
• with missing gender (SEX=missing), age (AGE=missing), quarter (DQTR=missing), year (YEAR=missing), or principal diagnosis (DX1=missing)

_____
1. Only for cases that otherwise qualify for the numerator.
2. If day of procedure is not available in the input data file, the rate may be slightly lower than if the information were available.

De.1. **Measure Type:** Outcome

S.23. **Data Source:** Administrative claims
S.26. **Level of Analysis:** Facility

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

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

Summary of evidence:

- The developers conducted an environmental scan to identify studies relevant to the outcome of interest. Several studies have examined the scientific acceptability of the PSI09 measure. These studies have demonstrated moderate to high positive and negative predicative values. They also present results from several studies that demonstrate that perioperative hemorrhage is preventable.

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*

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

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

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

- The developer provides information on a reference population rate and distribution of hospital performance for the measure between 2011-2013. The reference population is limited to states with present on admission data (POA).
- Between 2011-2012 the mean rate per 1000 surgical discharges was 3.432 (n=11,0043,343) and between 2012-2013 the mean rate was 3.613 per 1000 surgical discharges (n=10,780,407).
- The reference population is limited to states with POA data. The developer did not include data from prior to 2011 because many states were not report POA data before 2011.

Disparities

- The developer provides rates stratified by gender, age, payer, race/ethnicity and residence. The rates vary by these characteristics but there is no indication of whether or not these differences are significant.

Questions for the Committee:

- *Specific question on information provided for gap in care.*
- *Is there a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*

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)

1. Importance to Measure and Report

1a. Evidence to Support Measure Focus

Comments: ** This is an outcome measure. Developer did an environmental scan to identify relevant studies. Several studies of the accepability of the PS109 measure have shown moderate to high positive and negative predictive values. Several studies were also presented that show that perioperative hemorrhage is preventable.

** linked to outcomes and discussed preventive items

1b. Performance Gap

Comments: ** "2011-13 data are presented on a reference population rate an distribution of hospital performance. 2011-12 mean rate was 3.432; 2012-13 mean rate was 3.613. Rates are stratified by gender, age, pyer, race/ethnicity and residence and vary by these characteristics, but not indication is provided on their significance.

** developers provide gaps in care and tested for disparities

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Administrative claims

Specifications:

- This measure assesses the rate of perioperative hemorrhage or hematoma cases involving a procedure to treat the hemorrhage or hematoma in patients age 18 or older.
- The level of analysis for this measure is the hospital/acute care facility.
- The time period for data is either one or two years for users with a complete sample of hospital discharges; the developer notes that the signal variance parameters in the software assume at least a one-year time period, and that users may use longer time periods if desired.
- The denominator population for this measure (surgical discharges) is defined using ICD-9/10 operating room procedure codes and Medicare Severity Diagnosis-Related Group (MS-DRG) codes.
- The numerator identifies discharges with any secondary ICD-9-CM or ICD-10-CM diagnosis codes for perioperative hemorrhage or hematoma and any-listed ICD-9-CM or ICD-10-PCS procedure codes for treatment of hemorrhage or hematoma.
- The measure is expressed as a risk-adjusted rate per 1,000 surgical discharges; the risk-adjusted rate is computed using indirect standardization as the observed rate divided by the expected rate, multiplied by the reference population rate.

Questions for the Committee :

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

SUMMARY OF TESTING

Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**
Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing

- To demonstrate reliability, the developer conducted a signal-to-noise analysis of the measure score.
- The developer assessed the measure's signal-to-noise ratio by comparing the degree to which risk adjusted rates differ across hospitals (the signal) to the degree of precision of the rates within hospitals (the noise):
    - The signal-to-noise ratio was calculated at the hospital level and then summarized across the entire population of US hospitals.
- The developer notes that hospital size has an impact on reliability, and that smaller hospitals have less reliable rates due to very small denominators (the number of patients at risk). For this reason, the overall signal-to-noise ratio for the measure is calculated as a weighted estimate, using a method that reduces the influence of smaller hospitals.

Results of reliability testing

- To report the results of reliability testing, the developer grouped hospitals into deciles by size, and provided the average signal-to-noise ratio for each decile, as well as an overall reliability score:

Table 2. Signal-to-Noise Ratio by Hospital Size Decile, PSI 09 Perioperative Hemorrhage or Hematoma Rate Using 2-year pooled data (2012-2013)

Hospital Size Decile	Number of Hospitals	Avg. Number of Discharges per Hospital in Decile	Avg. Signal-to-Noise Ratio for Hospitals in Decile
1 (smallest)	357	43.9	0.0342
2	357	219.7	0.0893
3	358	521.4	0.1822
4	357	957.2	0.2741
5	357	1,533.5	0.3796
6	358	2,257.4	0.4778
7	357	3,131.7	0.5630
8	358	4,308.4	0.6456
9	357	6,202.5	0.7349
10 (largest)	357	11,001.4	0.8399
Overall	3,573	3,017.2	0.6661

- The developer observes that signal-to-noise ratios were smaller for hospitals with fewer than 1,534 qualifying discharges per year (average signal-to-noise ratio less than 0.38). For this reason, the developer recommends the use of 'smoothed rates', which bring scores toward the mean, particularly for smaller hospitals.
    - The developer notes that smoothed rates are implemented in the AHRQ software.
- The developer argues that there is no universally accepted threshold of "adequate" signal to noise ratio, stating that:
    - Different methods of calculating reliability and signal-to-noise (e.g., split sample or test-retest reliability of the data, different methods of calculating the hospital signal-to-noise ratio) result in different distributions of reliability scores. In addition, "adequate" depends on the specific application and judgment of the user. For instance, if a complication such as mortality is very important (e.g. leads to great harm to the patient) a lower reliability may be acceptable. However, the AHRQ QI program generally considers ratios between 0.4 – 0.8 as acceptable. It is rare to achieve reliability above 0.8, using hospital signal-to-noise ratios as an indicator of reliability.
- The developer considers the overall signal-to-noise ratio for this indicator to be good at 0.67.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *What do you think about the developer's findings on the reliability of smaller hospitals and the developer's approach to addressing these issues?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

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

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ Yes ☐ Somewhat ☐ No

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

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.

SUMMARY OF TESTING

Validity testing level ☐ Measure score ☐ Data element testing against a gold standard ☒ Both

Method of validity testing of the measure score:

- ☐ **Face validity only**
- ☒ **Empirical validity testing of the measure score**

Validity testing method:

- To demonstrate validity, the developer provides several methods of testing:
  - Data element validity
    - The developer notes that a number of validity assessments of PSI 09, particularly of the validity of the PSI 09 algorithm and administrative data to capture postoperative hemorrhage or hematoma, have been published in the peer review literature; the developer summarizes the most relevant of these studies.
  - Measure score validity
    - the developer utilized a structured panel review to evaluate face validity of the measure from a clinical perspective.
    - 7 members of a multispecialty panel and 6 members of a surgical subspecialty panel completed a 10-item questionnaire, discussed the measure on a moderated conference call, then completed the questionnaire again to provide their final ratings.

Validity testing results:

- Data element validity
  - Early studies of the validity of the PSI 09 algorithm for identifying postoperative hemorrhage and hematoma (prior to the implementation of present-on-admission flags) found moderate-to-high positive predictive values (PPV); studies conducted after AHRQ expanded the list of procedure codes that qualify as treatment for hemorrhage or hematoma have seen improved sensitivity and PPV of the measure.
- Face Validity

- the developer reports that in their clinical panel review, the indicator had high face validity for use in quality improvement and hospital comparative assessments.

Questions for the Committee:

- *Does the information provided by the developer demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

[Summarize and analysis of exclusions]

- The measure excludes the following cases:
  - with a principal ICD-9-CM or ICD-10-CM diagnosis code (or secondary diagnosis present on admission(1) for perioperative hemorrhage or postoperative hematoma
  - where the only operating room procedure is for treatment of perioperative hemorrhage or hematoma
  - with any secondary ICD-9-CM or ICD-10-CM diagnosis codes for perioperative hemorrhage or hematoma and any-listed ICD-9-CM or ICD-10-PCS procedure codes for treatment of perioperative hemorrhage or hematoma occurring before the first operating room procedure(2)
  - with any-listed ICD-9-CM or ICD-10-CM diagnosis codes for coagulation disorder
  - MDC 14 (pregnancy, childbirth, and puerperium)
  - with missing gender (SEX=missing), age (AGE=missing), quarter (DQTR=missing), year (YEAR=missing), or principal diagnosis (DX1=missing)
- Using 2013 data from 34 states, the developer examined the percent of potential denominator cases excluded by each criterion as listed in the measure specifications.
- The results of that analysis are provided in a table in the submission form.
- The developer notes that the intent of this measure is to isolate those hemorrhages that can truly be linked to a surgical procedure and are of sufficient severity to be consequential to the patient. The exclusions are designed to minimize false positive events; the developer states that the empirical analysis supports these exclusions, as they capture a non-trivial number of events.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☐ **None** ☒ **Statistical model** ☐ **Stratification**

Conceptual rationale for SDS factors included ? ☒ **Yes** ☐ **No**

SDS factors included in risk model? ☐ **Yes** ☒ **No**

Risk adjustment summary

- The measure is expressed as a risk-adjusted rate per 1,000 surgical discharges; the risk-adjusted rate is computed using indirect standardization as the observed rate divided by the expected rate, multiplied by the reference population rate.
  - The observed rate is the number of discharge records where the patient experienced the PSI adverse event divided by the number of discharge records at risk for the event.
  - The expected rate is computed as the sum of the predicted value for each case divided by the number of cases for the unit of analysis of interest (i.e., hospital). The risk adjusted rate is computed using indirect

standardization as the observed rate divided by the expected rate, multiplied by the reference population rate.
    - The predicted value for each case is computed using a hierarchical model (logistic regression with hospital random effect) and covariates for gender, age (in 5-year age groups, except for the youngest age range), Modified Diagnosis Related Groups, which are the base MS DRGs without any distinction for "comorbidity and complications" (CC/MCC), AHRQ Comorbidity Index, Major Diagnosis Categories (MDC) based on the principal diagnosis, and transfer in from another acute care hospital.

- Clinical Risk Adjustment
    - To select the risk-adjustment factors, the developer considered a standard set of covariates grouped into four categories: demographics, severity of illness, comorbidities and transfer-in status. Covariates that were considered as potential risk adjusters included gender and age, MDC, Modified Diagnostic Related Groups (MDRGs) (defined as the base MS-DRG without comorbidity or complication distinctions), AHRQ Comorbidity Software categories and whether they were transferred from another facility. Only those covariates with at least 30 cases for PSI 09 are retained. A parsimonious model was identified using backward stepwise selection with bootstrapping.
    - The measure's risk model includes ## risk categories, including 26 age-gender categories in 5-year age categories between ages 30 and 89, and 2 age-gender categories ranging from below age 30 (i.e. 18-29) as one category and ages 90+ as another category, transfer in from another acute care facility and 17 comorbidities.
    - The remainder of selected risk factors account for the reason for admission and the type of surgery that was performed during the hospitalization, including MDC and MS-DRGs collapsed to remove Complication or Comorbidity/ Major Complication or Comorbidity (CC/MCC) distinctions.
    - Additional details on the risk adjustment approach are available in the submission form and in supplemental materials provided by the developer.
    - To validate their risk-adjustment approach, the developer conducted an analysis to evaluate how strongly the risk adjustment model is associated with the event of interest.
        - The measure of discrimination, how well the risk adjustment model distinguishes events from non-events, is the c-statistic (also known as the area under a receiver operating characteristic curve)
            - The c-statistic is a measure of the proportion of discordant pairs of observations for which the observation with the event had a higher predicted probability from the risk-adjustment model than the observation without the event
        - The developer also evaluated the calibration of the risk adjustment model by evaluating how closely observed and predicted rates compare across deciles of the predicted rate. A model that is well calibrated will have observed values similar to predicted values across the predicted value deciles.
        - The results of this analysis are provided in a table in the submission form.
    - The developer's interpretation of their analysis is that the risk-adjustment model has moderately high discrimination, based on a **c-statistic of 0.7689** (i.e., in 77% of randomly selected pairs of discordant observations, the patient who experienced PSI 09 had a higher probability of experiencing the event than the patient who did not).
    - The developer also suggests that the measure is well calibrated, as the **observed to predicted ratio values across the deciles range between 0.87 to 1.09** for all deciles except the lowest decile

- SDS Adjustment
    - The developer observes that empirical studies have generally demonstrated minimal differences in PSI 09 rates across racial/ethnic categories, and contends that there is no evidence or causal model to suggest that SDS factors are associated with perioperative hemorrhage or hematoma independent of quality of care, or are mediated by pre-hospital care (which may not fall within the proper realm of hospital accountability).
        - Accordingly, consistent with the guidance provided by NQF in the SDS Trial Period FAQs, AHRQ believes that it would be inappropriate to include other SDS variables in the risk-adjustment

approach for PSI 09, which is an in-hospital outcome measure.

Questions for the Committee:

- *Is an appropriate risk-adjustment strategy included in the measure?*
- *Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented?*
- *Are all of the risk adjustment variables present at the start of care? If not, describe the rationale provided.*
- *Do you agree with the developer's rationale that there is no conceptual basis for adjusting this measure for SDS factors?*
- *Do you agree with the developer's decision, based on their analysis, to not include SDS factors in their risk-adjustment model?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- To determine if meaningful differences in performance measure scores among measured entities can be identified, the developer assesses the probability that a hospital is higher or lower than a benchmark or threshold, given hospital size.
- The developer suggests this analysis reflects whether the indicator can discriminate the best performing hospitals from the lower performing hospitals.
- The developer provides a table reporting the proportion of hospitals above (better than) and below (worse than) the Benchmark and Threshold rates and the proportion not classified as either above or below. The proportion of hospitals not classified as either better or worse have rates that fall within the 95% confidence interval.
- The developer's interpretation of their analysis is that, over all hospitals, this indicator has modest discrimination for identifying low or high performing hospitals; 35% of hospitals can be classified as better or worse than the threshold (the percentage classified as either above or below the threshold) and 33% better or worse than the benchmark (the percentage classified as either above or below the benchmark).
- The developer also notes that discrimination increases as hospital size increases.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

N/A

2b7. Missing Data

- With regard to missing data, the developer reports that PSI 09 excludes cases with missing discharge disposition, age, sex, discharge quarter, discharge year, and principal diagnosis. These variables are required for indicator construction and are required of all hospital discharge records.
- The developer notes that for these variables, frequencies of missing data are typically less than 1% of the state database, suggesting it is unlikely the bias would occur from such a low frequency of missing data.
- The developer concludes that exclusion of cases with missing data for these key variables is appropriate.

Guidance from the Validity Algorithm

[Box 1] Specifications consistent with evidence → [Box 2] Potential threats to validity addressed → [Box 3] Empirical validity testing conducted using the measure as specified → [Box 6] Testing NOT conducted at the measure score level → [Box 10] Testing conducted with patient-level data elements → [Box 11] Method described and appropriate → [Box 12] High or moderate certainty or confidence that measure scores are valid → [Box 12b]

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

Committee pre-evaluation comments

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

2. Scientific Acceptability of Measure Properties

2a1. & 2b1. Specifications

Comments: ** Specifications are clear and consistent with the evidence.

** "provide validity and reliability. signal to noise OK but not great. question is still how reliable using administrative data and can surgeon game the system

2a2. Reliability Testing

Comments: ** Measure score reliability testing was done using signal-to-noise analysis. Hospital size has an impact on reliability; therefore, the developer recommends using smoothed rates. The overall signal-to-noise ratio is 0.67 = good

** provide testing. agree it is moderate

2b2. Validity Testing

Comments: ** "Both measure score and data element validity testing were done using several methods, including studies of PSI 09 validity and a panel of medical and surgical experts.

** testing was adequate

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: ** Exclusions are supported. A risk-adjustment model is included that has moderately high discrimination (c-statistic of 0.7689). Measure has overall modest discrimination for identifying low or high performing hospitals. There is a very low level of missing data (<1%).

** Exclusions are clearly stated. measure can indicate difference in quality. need to discuss if administrative data truly linked to significant bleeding

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.

- This measure is generated or collected by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)
- ALL data elements are in defined fields in electronic claims.
- Because the indicator is based on readily available administrative billing and claims data, feasibility is not an issue.
- This version of the indicator requires present-on-admission (POA) data for risk-adjustment and for specification of the numerator and denominator.
- POA indicators were added as data elements to the uniform bill form (UB-04) effective October 1, 2007. Hospitals incurred a payment penalty for not including POA status on Medicare records beginning October 1, 2008. Each of the secondary diagnoses in a discharge record can be flagged as "present at the time the order for inpatient admission occurs" or not.

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

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: ** All data elements are in electronic claims.
** no concerns except potential for gaming

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

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

Current uses of the measure [from OPUS]
Publicly reported? ☒ **Yes** ☐ **No**

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

Accountability program details:

- Arizona Department of Health Services, AZ Hospital Compare, MONAHRQ website. Hospital quality ratings from all hospitals in Arizona.
- CareChex (Division of Quantros), Provides comprehensive reports of hospitals to consumers, providers and purchasers.
- Cigna - Centers of Excellence Hospital Value Tool – Health insurance company
- Connecticut Department of Health Services, CT Hospital Compare, MONAHRQ website - Hospital quality ratings from all hospitals in Connecticut
- Healthcare Association of New York State - Supports availability of hospital quality and safety information to help patients make choices and assist providers in improving care
- HealthGrades - Healthgrades measures 40 million patient records from 4,500 hospitals nationwide for the most recent three-year period. Consumer-targeted hospital and provider ratings
- Illinois State Government - Illinois Hospital Report Card and Consumer Guide to Health Care
- Iowa Healthcare Collaborative - Hospital quality ratings from hospitals in Iowa
- Kentucky Hospital Association Quality Data - Hospital quality ratings from most hospitals in Kentucky
- Maine Health Data Organization (MHDO), MONAHRQ Website - Hospital quality ratings from all hospitals in Maine
- Nevada Compare Care, MONAHRQ website - Hospital quality ratings from most hospitals in Nevada
- Nevada Hospital Association - Transparency and Performance of Nevada hospitals for specific clinical indicators
- New Jersey Department of Health - Public report of PSI performance for New Jersey Hospital
- Niagara Health Quality Coalition, New York State Hospital Report Card - Consumer focused public report of quality indicator performance for NY hospitals.
- Norton Healthcare -Report patient satisfaction scores in Norton Healthcare hospitals and their performance on nationally recognized quality indicators
- Oklahoma State Department of Health, MONAHRQ - Compares quality ratings on hospitals across Oklahoma
- Texas Health Resources - Provides quality and safety reports for all Texas Health Resources
- U.S. News and World Report - National publication that lists ratings of U.S. medical centers based on performance

- Virginia Health Information - Compares quality ratings on hospitals across Virginia
- Washington State, MONAHRQ website - Information system of inpatient care utilization, quality, and potentially avoidable stays in Washington State's community hospitals
- WHA Information Center (Wisconsin Hospital Association) - Wisconsin Inpatient Hospital Quality Indicators Report
- Quality Improvement (Internal to the specific organization) - Greenville Health System - Data collected from four hospitals in Greenville Health System, compared with internal rates

Improvement results:

- Rates of this measure have decreased slightly during 2011-2013 from 4.9 to 4.4 cases/1,000 hospitalizations. This may reflect improvements in care or motivation of providers to adjust documentation and coding practices to minimize the use of the perioperative hemorrhage and hematoma diagnosis codes.

Unexpected findings (positive or negative) during implementation:

- No evidence has been identified suggesting unintended consequences for this measure.

Potential harms:

- No evidence has been identified suggesting potential harm for this measure.

Feedback :

- Developer did not identify any specific feedback loops related to this 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

4. Usability and Use

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: ** The measure is currently publicly reported and used in accountability programs
** if PPV and NPV is correct and gaming can be minimized then canbe used for public reporting

Criterion 5: Related and Competing Measures

Related or competing measures

- N/A

Harmonization

- N/A

Pre-meeting public and member comments

Submitted by: Armstrong Institute for Patient Safety and Quality at Johns Hopkins University

We support efforts to measure patient safety in hospitals. We believe that valid and reliable measures of patient safety events are the foundation to improving performance and holding hospitals accountable.

Given the recent article by Winters et al. in Medical Care that found this measure did not meet validity thresholds when measured against the reference standard of a medical chart review, we would urge the standing committee to review the Medical Care article as part of their careful evaluation of the measure's validity.

Winters BD, Bharmal A, Wilson RF, Zhang A, Engineer L, Defoe D, Bass EB, Dy S, Pronovost PJ. Validity of the Agency for Health Care Research and Quality Patient Safety Indicators and the Centers for Medicare and Medicaid Hospital-acquired Conditions: A Systematic Review and Meta-Analysis. Medical care. 2016 Apr.

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 2909

Measure Title: Perioperative Hemorrhage or Hematoma Rate (PSI 09)

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: 0531 Patient Safety Composite for Selected Indicators (PSI 90)

Date of Submission: 5/13/2016

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

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

Outcome

☒ Health outcome: Perioperative Hemorrhage or Hematoma

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

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

☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Clinically, this indicator is intended to capture preventable and significant perioperative hemorrhage or hematoma events that are in excess of what is expected for the surgery type. The intent is to capture bleeding-related events that are severe or involve a delay in diagnosis or treatment requiring reoperation, as these events are associated with a significant increase in risk to the patient. Such events are often associated with the technical skill and judgment of the surgeon, especially when the hemorrhage is not recognized during the initial procedure and requires reoperation on a subsequent day. Best practices to prevent perioperative hemorrhage and hematoma include taking steps to address and avoid technical errors such as inadequate ligation, cauterization, clipping, or stapling of blood vessels; failure to recognize transection of a minor vessel; or defects in vascular anastomoses. Additional patient management processes that can contribute to PSI09 events include excessive anticoagulation; inadequate correction or reversal of coagulopathy; failure to replace clotting factors in cases involving large-volume blood loss; and intraoperative hypothermia.

Surgical textbooks such as Mulholland and Doherty's *Complications in surgery* (Philadelphia, PA: Lippincott, Williams & Wilkins, 2011) and professional guidelines address many of these issues.[1] For example, the World Health Organization (WHO) Guidelines for Safe Surgery state on p. 36:

"The first step in mitigating blood loss during an operation is prevention. Known coagulation deficits should be corrected before surgery whenever clinically possible. The surgical, anaesthetic and nursing personnel involved in an operation should all be aware of the potential for major blood loss before the procedure and be prepared for it…"[2]

The Practice Guidelines for Perioperative Blood Management by the American Society of Anesthesiologists focus on the use of anti-fibrinolytic pharmacologic therapies to prevent or minimize perioperative hemorrhage, especially in the setting of cardiopulmonary bypass:

"**ε-Aminocaproic Acid**. Meta-analysis of placebo-controlled RCTs indicate that the use of ε-aminocaproic acid administered before and/or during a procedure is effective in reducing total perioperative blood loss and the number of patients transfused in major cardiac, orthopedic, or liver surgery (Category

A1-B evidence); equivocal findings are reported for the volume of blood transfused (Category A1-E evidence). An RCT comparing ε-aminocaproic acid with placebo reports less blood loss and lower RBC transfusion requirements when ε-aminocaproic acid is administered for prophylaxis of excessive bleeding after total knee replacement surgery and before tourniquet deflation (Category A3-B evidence).

Tranexamic Acid. Meta-analysis of placebo-controlled RCTs indicate that tranexamic acid for prophylaxis of excessive bleeding administered before and/or during a procedure is effective in reducing perioperative blood loss, the number of patients transfused, and the volume of blood products transfused (CategoryA1-B evidence). Randomized trials comparing tranexamic acid with placebo or no tranexamic acid controls report no differences for stroke, myocardial infarction, renal failure, reoperation for bleeding, or mortality (Category A2-B evidence).

Meta-analysis of placebo-controlled RCTs indicate that tranexamic acid for prophylaxis of excessive bleeding initiated after a knee and hip arthroplasty and before tourniquet deflation compared with placebo also reported lower blood loss volumes (Category A1-B evidence). One RCT did not show efficacy when tranexamic acid was administered after cardiac surgery and continued for 12 h (Category A3-E evidence).

Survey Findings: The consultants and ASA members both agree…

- Use antifibrinolytic therapy for prophylaxis of the use of allogeneic blood transfusion in patients undergoing cardiopulmonary bypass.
- Consider using antifibrinolytic therapy for prophylaxis in certain orthopedic procedures such as knee replacement surgery.
- Consider using antifibrinolytic therapy for prophylaxis in liver surgery and other clinical circumstances at high risk for excessive bleeding."[3]

The same guidelines also address preoperative discontinuation of anticoagulants and antiplatelet agents when the risks exceed the benefits, perioperative reversal of anticoagulants, and intraoperative monitoring for blood loss and coagulopathy. Many relevant processes of care are also outlined in the chapter in the AHRQ toolkit that targets the prevention of PSI09 events. For example, this document suggests that "proper management of blood loss, including frequent dressing checks, is key to management of postoperative hemorrhage and hematoma."[4]

A recent Cochrane review assessed the comparative effects of three anti-fibrinolytic drugs (aprotinin, tranexamic acid (TXA), and epsilon aminocaproic acid (EACA)) on blood loss during surgery, the need for red blood cell (RBC) transfusion, and other adverse events. They reported that "aprotinin reduced the need for reoperation due to bleeding by a relative 54% (RR 0.46, 95% CI 0.34 to 0.62). This translates into an absolute risk reduction of 2% and a number needed-to-treat (NNT) of 50 (95% CI 33 to 100). A similar trend was seen with EACA (RR 0.32, 95% CI 0.11 to 0.99) but not TXA (RR 0.80, 95% CI 0.55 to 1.17)."[5] Spertus et al. (2015) used percutaneous coronary intervention data from 9 US hospitals to compare the use of bleeding avoidance strategies and bleeding rates before and after implementation of a validated risk model to determine individual patient risk of bleeding [developed by the American College of Cardiology's National Cardiovascular Data Registry (NCDR) Catheterization PCI Registry]. They compared 7408 pre-intervention procedures with 3529 post-intervention procedures and found that the use of the risk stratification protocol was also associated with lower bleeding rates compared to non-interventional sites (1.0% v 1.7%; odds ratio 0.56, 0.40 to 0.78; 0.62, 0.44 to 0.87), after adjustment.[6]

A limited number of older studies evaluated the actual occurrence of process failures in association with PSI 09 events. In a case control study involving 1,025 Medicare discharges from acute-care hospitals in California and Connecticut in 1994, nurse-identified process of care failures were relatively frequent among major surgical cases with postprocedural hemorrhage or hematoma (29/44=66%), after excluding patients who had hemorrhage or hematoma at admission.[7] Specifically, "problems with technical care during a procedure were present in 12 of 17 surgical… cases of postprocedural hemorrhage or hematoma".[8] Physician reviewers identified potential quality problems in 37% of major surgery patients with this event, versus 2% of unflagged controls.[8] However, cases flagged on this indicator and unflagged controls did not differ significantly on a composite of 17 generic process criteria, confirming previous findings in elderly Medicare beneficiaries from Massachusetts, Alabama, Iowa, and New York.[9]

Finally, the incidence of patient safety events can be influenced by certain health system characteristics such as provider-to-patient ratios, provider training, involvement of physicians-in-training, and staffing hour regulations. However, studies examining the impact of health system characteristics such as teaching status, safety climate, bed size, and nurse staffing hours have been inconclusive.[10-13] **Before mandatory present on admission (POA) reporting, rates were significantly higher at major teaching hospitals than at nonteaching hospitals in the Nationwide Inpatient Sample (OR 1.20 [95% CI 1.01 to 1.42]), but not in the**

Veterans Health Administration.[10, 11] Chen et al. analyzed rates of PSI09 (version 3.1a) among veteran dual users (i.e., those with hospitalizations in both the VA and the private sector with Medicare coverage) during 2002 to 2007 and found the risk-adjusted rate of PSI09 in the VA (3.3; 95% CI 3.0-3.6) to be significantly higher than in the private sector (2.1; 95% CI 1.9-2.4); dual users hospitalized in the VA had 1.73 times higher odds of PSI 09 than those hospitalized in the private sector (95% CI 1.48-2.03).[12] Rivard et al. (2010) examined over 4500 responses to the Patient Safety Climate in Healthcare Organizations survey **and found that the PSI09 rate was not significantly associated with any of the 11 dimensions of patient safety culture, adjusting for major teaching status, metropolitan area, and nurse-staffing ratio (p>0.10 for all comparisons).**[10] A study using the national inpatient data from the Japanese Diagnosis Procedure Combination database reported postoperative bleeding and perforation in 331 (4.4%) and 13 patients (0.2%) who underwent colorectal endoscopic submucosal dissections (n=7567). "Multivariable logistic regression analysis showed that the very high hospital volume group had a significantly lower proportion of severe postoperative bleeding than the very low hospital volume group (OR = 0.48 [95 % CI, 0.27-0.83]; p = 0.009)".[14]

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☒ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

Please note that this is an outcome measure, so a systematic review of the body of evidence that supports the performance measure is not required. However, information is provided in 1a.8 below, to provide additional context and support for the measure.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

Not applicable

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**. Not applicable

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

Not applicable

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

Not applicable

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

Not applicable

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

Not applicable

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

Not applicable

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

Not applicable

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

Not applicable

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

Not applicable

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Not applicable

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

Not applicable

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Not applicable

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

Not applicable

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

Not applicable

1a.7.2. Grade assigned for the quoted evidence with definition of the grade:

Not applicable

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

Not applicable

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: Click here to enter date range

Not applicable

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

Not applicable

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

Not applicable

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

Not applicable

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

Not applicable

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for each new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

Not applicable

1a.8 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.8.1 What process was used to identify the evidence?

Formal environmental scans of the literature, including routine PubMed searches are performed to continually update evidence. The current evidence review results presented below constitute the most recent update, conducted in January 2016. Search terms included relevant MeSH terms (hematoma, hemorrhage, hypovolemic shock, postoperative,

perioperative, or surgical complications). We combined this clinical search string with MeSH terms (hospitals, patient admission, inpatient, indicator, epidemiol*, statistic, patient safety, AHRQ, prevalence, incidence or utilization) to identify studies examining quality of inpatient care. The search was limited to English-language publications. For completeness we also tested more inclusive search strings. Below we have provided a summary of the most up-to-date evidence.

1a.8.2. Provide the citation and summary for each piece of evidence.

Scientific Acceptability: Reliability/Validity

Several studies have examined the scientific acceptability of PSI09, as measured by its positive predictive value (PPV). **Utter and colleagues (2013) evaluated PSI09 (modified version 4.0a) by examining a sample of indicator positive and negative records from 35 hospitals between 2006 and 2009. They found that of 181 indicator-positive records, 168 (93%) involved a correctly coded bleeding event that occurred during the same hospitalization (true positive from a coding perspective), yielding a PPV of 95% (95% CI, 90–98%). Of the 13 false-positive records, 11 had a hemorrhage or hematoma present on admission (POA) and two had no hemorrhage or hematoma. However, only 126 of the 181 cases experienced a true postoperative hemorrhage or hematoma following an index operation that required a subsequent procedure to treat the complication, yielding a PPV from the clinical perspective of 78% (95% CI, 58–90%). In a separate sample of 281 indicator-negative records, 32 were false negatives, yielding a sensitivity of 42% (95% CI 23-64%) and negative predictive value (NPV) of 99.7% (95% CI 99.0 to 99.9%), after accounting for the sampling design.**[15]

A similar study by Borzecki et al. (2011) examined the PPV of PSI09 (version 3.1a) using administrative data from the VA from fiscal years 2003-2007. In total, 1,998 records were flagged as meeting the criteria for PSI09. Of these, 112 hospitalizations from a total of 28 hospitals were randomly selected for abstraction. As in the Utter study, trained nurses conducted a detailed review of the 112 records, confirming a clinical diagnosis of postoperative hemorrhage or hematoma in 84 (PPV=75%; 95% CI 66% to 83%). Of the records examined, false positives were due to coding inaccuracies (4 cases), hemorrhages or hematomas that were present on admission (8 cases), "intraoperative bleeding controlled during the original procedure without subsequent bleeding or need for management" (6 cases), postoperative hemorrhage or hematoma that did not actually require a separate procedure to treat it (5 cases), and hematoma following a non-eligible index procedure, such as a diagnostic cardiac catheterization (3 cases). The authors concluded that PSI09 could be improved by improving coding practices and implementing "present on admission"(POA) coding.[16]

These studies were recently meta-analyzed by Winters et al.,[17] who reported a pooled PPV estimate of 79% (95% CI, 73-84%), but this estimate is incorrect because the authors double-counted the same VA study, failed to account for POA reporting, and failed to include more recent studies cited below. These two studies were performed prior to the widespread implementation of POA flags, required by the Centers for Medicare & Medicaid Services starting in 2008. Use of POA flags should further improve the performance of PSI09, as supported by Rosen's revised PPV estimate of 82% (95% CI 73-89%) and a single site study by Ramanathan et al. (2013), who reported a PPV for PSI09 of 97% (95% CI 85-100%).[18,19]

To improve the sensitivity of PSI09, AHRQ expanded the list of procedure codes that qualify as treatment for hemorrhage or hematoma (version 4.4). This expanded list was based on detailed review of ICD-9-CM to identify codes for control of hemorrhage, drainage (of hematoma), embolization, evacuation of a pelvic clot, ligation or suture of a blood vessel, exploration of a space in which bleeding may occur, and endoscopy (for diagnosis or treatment of hemorrhage). The expanded list was tested in the same sample by Utter et al. (2013) who reported that the updated specification for PSI09 had a sensitivity of 85% (95% CI, 67-94%) and a PPV of 76% (95% CI, 60-88%).[15] Based on user feedback, additional revisions to the qualifying procedure list are being tested to further improve PPV while maintaining high sensitivity.

AHRQ renamed the indicator to "Perioperative Hemorrhage or Hematoma" (from "Postoperative Hemorrhage or Hematoma") in 2015 because the available ICD-9-CM diagnosis codes do not distinguish between intraoperative and postoperative hemorrhage events. AHRQ believes this distinction becomes possible again with ICD-10-CM coded data. Once AHRQ confirms that the distinction is possible through validity testing, AHRQ may revert back to "Postoperative Hemorrhage or Hematoma" for the ICD-10-CM/PCS specifications.

Disparities

The distribution of surgical complications can differ due to patient or other demographic characteristics. In a 10-year study of patients undergoing thyroid or parathyroid surgery using the Nationwide Inpatient Sample (NIS), cases that had the following characteristics had a higher risk of neck hematoma: "aged 65 years and older (OR 1.8; 95% CI 1.4-2.1), male sex (OR 1.3; 95% CI 1.2-1.4), African-American race (OR 1.5; 95% CI 1.2-1.7), from the South (OR 1.3; 95% CI 1-1.4), comorbidity score of 3 or more (OR 2; 95% CI 1.6-2.6), history of alcohol abuse (OR 2.7; 95% CI 1.6-2.5), Grave's disease (OR 3;, 95% CI = 2.1-4.1), and substernal thyroidectomy (OR = 3.3, 95% CI = 2.8-3.9)".[20] Another study by Browne et al. (2014) revealed that among patients undergoing primary hip or knee arthroscopy Medicaid patients had increased rates hematoma or seroma compared to non-Medicaid patients.[21] Among the sample, there were 978 cases of hematoma or seroma among Medicaid patients (or 0.9% of the Medicaid sample) compared to 726 cases (0.7%) among non-Medicaid patients. In a study based on the 2011 NIS, Nwaogu et al, (2015) determined that the incidence of postmastectomy bleeding using ICD-9-CM procedure codes 85.34-85.48 for mastectomy, diagnosis codes 174.0-174.9 for breast cancer, and diagnostic codes 998.11 and 998.12 for hemorrhage and excluding males and those with a history of a coagulation disorder. Of the total of 7907 discharges meeting inclusion criteria; 201 had bleeding complications (2.54%), with 42 cases requiring reoperation. On multivariate analysis, the presence of CHF was a significant predictor of bleeding complications (odds ratio [95% confidence interval], 2.45, 95% CI [1.25-4.92], P = 0.009).[22]

Relationship to Other Outcomes

PSI09 events are associated with a number of important and significant patient harms such as increased postoperative infection, hypovolemic or hemorrhagic shock, reoperation, complications from blood transfusion (such as transfusion-related acute cardiac overload (TACO) and transfusion-related lung injury (TRALI)), mortality and resource use.[1,2,11,22-30]

Research has established associations between PSI09 and other outcomes, including hospital readmissions, costs, length of stay, and mortality.[18,23-26] Cases from the 2000 Nationwide Inpatient Sample that were flagged by this PSI had 3.0% excess mortality, 3.9 days of excess hospitalization, and $21,431 in excess hospital charges, relative to carefully matched controls that were not flagged.[25] This finding was confirmed in the Veterans Health Administration system, where cases that were flagged by this PSI in 2001 had 5.1-8.0% excess mortality, 3.9-4.7 days of excess hospitalization, and $7,863-10,012 in excess hospital costs, relative to propensity-matched or multivariable regression-adjusted controls that were eligible but not flagged.[11] In another study based on State Inpatient Databases from seven states that permit linkage of serial hospitalizations, PSI 09 was associated with risk ratios of 1.03 (NS) for inpatient death, 1.18 ($p<0.01$) for readmission within three months, and 1.10 (NS) for readmission within one month, after adjusting for age, gender, payer, comorbidities, specific surgical DRGs, and APR-DRG severity levels.[27] Similarly, in a multivariable analysis of Veterans Health Administration data, hospitalizations with PSI 09 were 60% more likely to result in a readmission within 30 days than eligible hospitalizations without PSI 09 (18.8% versus 11.3%; OR=1.60, 95% CI 1.40 to 1.83), after adjusting for age, sex, comorbidities, and other PSI events (Rosen et al., 2013).[23] Ramanathan et al. (2014) retroactively examined data on surgical patients hospitalized between 2011 and 2012 at a single academic medical center and found that hospitalizations with PSI09 (version 3.1) were associated with a mean hospital LOS of 22.1 days, 64.5% included an intensive care unit stay, and 3.2% died in hospital.[28]

Several other studies have focused on narrower clinical cohorts. In an analysis of patient-level Medicare claims data for patients undergoing any of 6 cancer resections in 2005-2009, Short et al. found that after adjusting for patient factors (age, sex, race, income), hospital factors (hospital volume, surgeon volume, surgeon specialty

designation, hospital resources, patient characteristics) and tumor factors (tumor stage, site), costs increased significantly in association with postoperative hemorrhage or hematoma for four of the six types of cancer resection patients ($p<0.001$).[24] Based on an analysis of the 501,908 hospitalizations involving a brain tumor in the NIS between 2002 and 2010, Rahman et al. (2013) found that patients with postoperative hemorrhage or hematoma had significantly longer length-of-stay (LOS) (13.1 days vs 6.5 days; $p < 0.0001$), on average, than patients without this complication.[26] In another NIS-based study limited to patients with breast cancer hospitalized for a mastectomy in 2011, Nwaogu et al. (2015) reported a 1.3 day increase in the mean length of stay ($P < 0.0001$), a $5495 increase in the mean cost per hospital stay ($P < 0.0001$), and a reoperation rate of 2.5% (42 of 201) associated with a bleeding complication (as defined by ICD-9-CM codes 998.11, 998.12, 39.98, and 86.04).[22] De la Garza-Ramos and colleagues (2016) estimated the incidence of in-hospital morbidity and mortality following surgery for malignant brain tumors using the NIS from 2002 to 2011; patients who had experienced a hemorrhage/hematoma complication (based on an expanded list of ICD-9-CM codes [998.1–998.13] compared to PSI09) had 3.3 times higher odds of mortality (95% CI 1.6–6.6) than those who did not experience that surgical complication.[29] Finally, Ang and colleagues (2015) used 2013 data from the Florida Agency for Health Care Administration to evaluate trauma mortality using the AHRQ PSIs. Of the 939 PSI09 events (version 4.5) in 50,596 trauma patients, there were 101 deaths. With an adjusted "failure to prevent" observed-to-expected ratio of 3.53, PSI09 had the strongest influence on trauma mortality among the 10 PSIs reviewed.[31]

References

1. Mulholland MW, Doherty GM. *Complications in Surgery.* Philadelphia, PA: Lippincott, Williams & Wilkins; 2011.
2. *WHO Guidelines for Safe Surgery 2009: Safe Surgery Saves Lives*. Geneva2009.
3. American Society of Anesthesiologists Task Force on Perioperative Blood M. Practice guidelines for perioperative blood management: an updated report by the American Society of Anesthesiologists Task Force on Perioperative Blood Management*. *Anesthesiology.* 2015;122(2):241-275.
4. http://www.ahrq.gov/sites/default/files/wysiwyg/professionals/systems/hospital/qitoolkit/d4g-postophemorrhage-bestpractices.pdf
5. Henry DA, Carless PA, Moxey AJ, et al. Anti-fibrinolytic use for minimising perioperative allogeneic blood transfusion. *Cochrane Database Syst Rev.* 2011(3):CD001886.
6. Spertus JA, Decker C, Gialde E, et al. Precision medicine to improve use of bleeding avoidance strategies and reduce bleeding in patients undergoing percutaneous coronary intervention: prospective cohort study before and after implementation of personalized bleeding risks. *BMJ (Clinical research ed.).* 2015;350:h1302.
7. Iezzoni LI, Davis RB, Palmer RH, et al. Does the Complications Screening Program flag cases with process of care problems? Using explicit criteria to judge processes. *Int J Qual Health Care.* 1999;11(2):107-118.
8. Weingart SN, Iezzoni LI, Davis RB, et al. Use of administrative data to find substandard care: validation of the complications screening program. *Med Care.* 2000;38(8):796-806.
9. Iezzoni LI, Lawthers A, Peterson L, McCarthy E, Palmer RH, Cahalane M. *Project to validate the Complications Screening Program: Health Care Financing Administration.* 1998. HCFA Contract 500-94-0055.
10. Rivard PE, Elixhauser A, Christiansen CL, Shibei Z, Rosen AK. Testing the association between patient safety indicators and hospital structural characteristics in VA and nonfederal hospitals. *Med Care Res Rev.* 2010;67(3):321-341.
11. Rivard PE, Christiansen CL, Zhao S, Elixhauser A, Rosen AK. Advances in Patient Safety: Is There an Association Between Patient Safety Indicators and Hospital Teaching Status? In: Henriksen K, Battles JB, Keyes MA, Grady ML, eds. *Advances in Patient Safety: New Directions and Alternative Approaches (Vol. 2: Culture and Redesign)*. Rockville (MD)2008.
12. Chen Q, Hanchate A, Shwartz M, et al. Comparison of the Agency for Healthcare Research and Quality Patient Safety Indicator Rates Among Veteran Dual Users. *Am J Med Qual.* 2013;29(4):335-343.
13. Rosen AK, Zhao S, Rivard P, et al. Tracking rates of Patient Safety Indicators over time: lessons from the Veterans Administration. *Med Care.* 2006;44(9):850-861.
14. Odagiri H, Yasunaga H, Matsui H, Fushimi K, Iizuka T, Kaise M. Hospital volume and the occurrence of bleeding and perforation after colorectal endoscopic submucosal dissection: analysis of a national administrative database in Japan. *Dis Colon Rectum.* 2015;58(6):597-603.
15. Utter GH, Zrelak PA, Baron R, et al. Detecting postoperative hemorrhage or hematoma from administrative data: the performance of the AHRQ Patient Safety Indicator. *Surgery.* 2013;154(5):1117-1125.

16. Borzecki AM, Kaafarani H, Cevasco M, et al. How valid is the AHRQ Patient Safety Indicator "postoperative hemorrhage or hematoma"? *J Am Coll Surg.* 2011;212(6):946-953 e941-942.
17. Winters BD, Bharmal A, Wilson RF, et al. Validity of the Agency for Health Care Research and Quality Patient Safety Indicators and the Centers for Medicare and Medicaid Hospital-acquired Conditions: A Systematic Review and Meta-Analysis. *Medical Care.* 2016.
18. Ramanathan R, Leavell P, Stockslager G, Mays C, Harvey D, Duane TM. Validity of Agency for Healthcare Research and Quality Patient Safety Indicators at an academic medical center. *Am Surg.* 2013;79(6):578-582.
19. Rosen AK, Itani KM, Cevasco M, et al. Validating the patient safety indicators in the Veterans Health Administration: do they accurately identify true safety events? *Medical care.* 2012;50(1):74-85.
20. Dehal A, Abbas A, Hussain F, Johna S. Risk factors for neck hematoma after thyroid or parathyroid surgery: ten-year analysis of the nationwide inpatient sample database. *The Permanente journal.* 2015;19(1):22-28.
21. Browne JA, Novicoff WM, D'Apuzzo MR. *Medicaid Payer Status Is Associated with In-Hospital Morbidity and Resource Utilization Following Primary Total Joint Arthroplasty.* Vol 962014.
22. Nwaogu IY, Bommarito K, Olsen MA, Margenthaler JA. Economic impact of bleeding complications after mastectomy. *The Journal of surgical research.* 2015;199(1):77-83.
23. Rosen AK, Loveland S, Shin M, et al. Examining the impact of the AHRQ Patient Safety Indicators (PSIs) on the Veterans Health Administration: the case of readmissions. *Med Care.* 2013;51(1):37-44.
24. Short MN, Aloia TA, Ho V. The influence of complications on the costs of complex cancer surgery. *Cancer.* 2014;120(7):1035-1041.
25. Zhan C, Miller MR. Administrative data based patient safety research: a critical review. *Qual Saf Health Care.* 2003;12 Suppl 2:ii58-63.
26. Rahman M, Neal D, Fargen KM, Hoh BL. Establishing standard performance measures for adult brain tumor patients: a Nationwide Inpatient Sample database study. *Neuro Oncol.* 2013;15(11):1580-1588.
27. Friedman B, Encinosa W, Jiang HJ, Mutter R. Do patient safety events increase readmissions? *Med Care.* 2009;47(5):583-590.
28. Ramanathan R, Leavell P, Wolfe LG, Duane TM. Agency for Healthcare Research and Quality patient safety indicators and mortality in surgical patients. *Am Surg.* 2014;80(8):801-804.
29. De la Garza-Ramos R, Kerezoudis P, Tamargo RJ, Brem H, Huang J, Bydon M. Surgical complications following malignant brain tumor surgery: An analysis of 2002-2011 data. *Clin Neurol Neurosurg.* 2016;140:6-10.
30. Brunicardi F, Andersen D, Billiar T. *Schwartz's Principles of Surgery, 10e.* Columbus, OH: McGraw-Hill Education; 2015.
31. Ang D, McKenney M, Norwood S, et al. Benchmarking statewide trauma mortality using Agency for Healthcare Research and Quality's patient safety indicators. *The Journal of surgical research.* 2015;198(1):34-40.

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
PSI09_Evidence_Form_160511_v3.docx

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., the benefits or improvements in quality envisioned by use of this measure)*
Clinically, this indicator is intended to capture preventable and significant perioperative hemorrhage or hematoma events that are in excess of what is expected for the surgery type. The intent is to capture the significant events, such as those that are severe or where there may be a delay in diagnosis or treatment requiring reoperation, as these events are associated with a significant

increase in harm to the patient.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
This table is also included in the supplemental materials.

Table 1. Reference Population Rate and Distribution of Hospital Performance for PSI 09 Perioperative Hemorrhage or Hematoma Rate in 2-year Pooled Data (2011-2013)

Overall Reference Population Rate

Year3	Number of Hospitals (Numerator)1	Outcome of Interest Population at Risk (Denominator)1	Observed Rate Per 1000 Surgical Discharges1
2011-2012	3,432	52,548 11,043,434	4.7583
2012-2013	3,613	48,663 10,780,407	4.5140

Distribution of Hospital-level Observed Rates in Reference Population

Year3	Number of Hospitals	Rates per 1000 Surgical Discharges (p=percentile)2						
		Mean	SD2	p5	p25	Median	p75	p95
2011-2012	3,432	3.81	3.83	0.00	1.56	3.52	5.21	8.90
2012-2013	3,613	3.64	4.13	0.00	1.32	3.28	5.05	8.50

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2011 - 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov (AHRQ QI Software Version 6.0)
1The observed rate refers to the total rate for all observations included in the reference population data (numerator) divided by the total combined eligible population of all hospitals included in the reference population data (denominator).
2The distribution of hospital rates reports the mean and standard deviation (SD) of the observed rates for all hospitals included in the dataset with at least one case in the denominator, as well as the observed rate for hospitals in the 5th, 25th, 50th (median), 75th, and 95th percentile. Standard deviation refers to the spread in observed values in relation to the mean.
3 Reference population is limited to states with present on admission data (POA). Since many states did not report POA data prior to 2011 we have not included testing prior to 2011.

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*
This table is also included in the supplemental materials.
Table 2. Perioperative Hemorrhage or Hematoma Rate (PSI 09) per 1,000 surgical discharges, by patient and hospital characteristics, 2013

Patient/hospital characteristic	Estimate	Std Error	p-value	(Ref Grp = *)	Lower 95% CL	Upper 95% CL
Total U.S.	4.37	0.03		4.32	4.43	
Patient Characteristics						
Age Groups:						
18-39*	3.98	0.08	*	3.82	4.14	
40-64	4.29	0.04	<.001	4.20	4.37	
65 and over	4.58	0.04	<.001	4.49	4.67	
Gender:						
Male*	4.63	0.04	*	4.55	4.71	

Female	4.14	0.04	<.001	4.06	4.22		
Patient Zip Code Median Income							
First quartile (lowest income)				4.28	0.11	0.128	4.07	4.49
Second quartile	4.21	0.07	0.006	4.08	4.34		
Third quartile	4.47	0.06	0.226	4.35	4.58		
Fourth quartile (highest income)*				4.41	0.04	*	4.33	4.49
Location of patient residence (NCHS):							
Rural	4.27	0.22	0.318	3.83	4.71		
Urban*	4.38	0.03	*	4.32	4.44		
Expected payment source:							
Private insurance*				4.29	0.05	*	4.19	4.38
Medicare1		4.38	0.04	0.070	4.30	4.46	
Medicaid		5.07	0.10	<.001	4.87	5.26	
Uninsured / self-pay / no charge				3.97	0.13	0.011	3.72	4.22
Other insurance	4.18	0.15	0.263	3.89	4.48		
Location of Care:							
Northeast*		4.41	0.07	*	4.27	4.55	
Midwest	4.24	0.06	0.036	4.13	4.36		
South	4.37	0.05	0.330	4.28	4.47		
West	4.52	0.06	0.118	4.40	4.64		

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov (AHRQ QI Software Version 6.0)
Rates are adjusted using the AHRQ QI PSI POA Reference Population for 2013 as the standard population. Age and gender are removed from models for the relevant strata.
NCHS - National Center for Health Statistics designation for urban-rural locations.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
N/A

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Patient/societal consequences of poor quality
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
This measure has been independently associated with increased mortality (3-8%), length of stay (3.9-4.7 days), and cost ($7,863-10,012), based on cohort studies involving non-federal acute care hospitalizations and hospitalizations in the Veterans Health Administration setting. Approximately 66% of records flagged for perioperative hemorrhage or hematoma demonstrate some deficiency in the process of care, with most of these representing intraoperative events.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
Rosen AK, Loveland S, Shin M, et al. Examining the impact of the AHRQ Patient Safety Indicators (PSIs) on the Veterans Health Administration: the case of readmissions. Med Care. 2013;51(1):37-44.
Zhan C, Miller MR. Administrative data based patient safety research: a critical review. Qual Saf Health Care. 2003;12 Suppl 2:ii58-63.
Iezzoni LI, Davis RB, Palmer RH, et al. Does the Complications Screening Program flag cases with process of care problems? Using explicit criteria to judge processes. Int J Qual Health Care. 1999;11(2):107-118.

HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2011-2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov (AHRQ QI Software Version 6.0)

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
Not applicable

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria 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):*
Surgery : Perioperative

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Safety

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.)*
http://1.usa.gov/1VVjaWz Note: The URL link will be updated for version 6.0 public release found via the module page: http://qualityindicators.ahrq.gov/Modules/psi_resources.aspx

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

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** PSI09_Technical_Specifications_160513.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
N/A

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Discharges, among cases meeting the inclusion and exclusion rules for the denominator, with:
• any secondary ICD-9-CM or ICD-10-CM diagnosis codes for perioperative hemorrhage or hematoma and any-listed ICD-9-CM or ICD-10-PCS procedure codes for treatment of hemorrhage or hematoma

Note that the ICD-10-CM specification is limited to postoperative hemorrhage or hematoma, whereas the ICD-9-CM specification captures both intraoperative and postoperative hemorrhage or hematoma (due to diagnosis codes that are less specific).

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
The time period is either one or two years for users with a complete sample of hospital discharges (i.e., "all payer" data). Note that the signal variance parameters in the software assume at least a one-year time period. Users may use longer time periods if desired.

S.6. **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, 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.
Please see attached excel file in S.2b. for version 6.0 specifications.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Surgical discharges, for patients ages 18 years and older, with any-listed ICD-9-CM or ICD-10-PCS procedure codes for an operating room procedure. Surgical discharges are defined by specific MS-DRG codes.

See Appendices: (attached in S.2b)
- Appendix A – Operating Room Procedure Codes
- Appendix E – Surgical Discharge MS-DRGs (for discharges on or after October 1, 2007)

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
Please see attached excel file in S.2b. for version 6.0 specifications.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Exclude cases:
• with a principal ICD-9-CM or ICD-10-CM diagnosis code (or secondary diagnosis present on admission(1) for perioperative hemorrhage or postoperative hematoma
• where the only operating room procedure is for treatment of perioperative hemorrhage or hematoma
• with any secondary ICD-9-CM or ICD-10-CM diagnosis codes for perioperative hemorrhage or hematoma and any-listed ICD-9-CM or ICD-10-PCS procedure codes for treatment of perioperative hemorrhage or hematoma occurring before the first operating room procedure(2)
• with any-listed ICD-9-CM or ICD-10-CM diagnosis codes for coagulation disorder
• MDC 14 (pregnancy, childbirth, and puerperium)
• with missing gender (SEX=missing), age (AGE=missing), quarter (DQTR=missing), year (YEAR=missing), or principal diagnosis (DX1=missing)

_____
1. Only for cases that otherwise qualify for the numerator.
2. If day of procedure is not available in the input data file, the rate may be slightly lower than if the information were available.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
Please see attached excel file in S.2b. for version 6.0 specifications.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
Not applicable

S.13. Risk Adjustment Type (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. Identify the statistical risk model method and variables *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
The predicted value for each case is computed using a hierarchical model (logistic regression with hospital random effect) and covariates for gender, age (in 5-year age groups, except for the youngest age range), Modified Diagnosis Related Groups, which are the base MS DRGs without any distinction for "comorbidity and complications" (CC/MCC), AHRQ Comorbidity Index, Major Diagnosis Categories (MDC) based on the principal diagnosis, and transfer in from another acute care hospital. A parsimonious model was identified using a backward stepwise selection procedure with bootstrapping. The expected rate is computed as the sum of the predicted value for each case divided by the number of cases for the unit of analysis of interest (i.e., hospital). The risk adjusted rate is computed using indirect standardization as the observed rate divided by the expected rate, multiplied by the reference population rate.

Additional information on methodology can be found in the Empirical Methods document on the AHRQ Quality Indicator website (www.qualityindicators.ahrq.gov). The Empirical Methods are also attached in the supplemental materials.

The specific covariates for this measure are as follows:

PARAMETER	LABEL
Intercept	Intercept
Sex | Age Demographics:	
M_AgeCat_1	Male | Age 18 - 29
M_AgeCat_2	Male | Age 30 - 34
M_AgeCat_3	Male | Age 35 - 39
M_AgeCat_4	Male | Age 40 - 44
M_AgeCat_5	Male | Age 45 - 49
M_AgeCat_6	Male | Age 50 - 54
M_AgeCat_7	Male | Age 55 - 59
M_AgeCat_8	Male | Age 60 - 64
M_AgeCat_9	Male | Age 65 - 69
M_AgeCat_10	Male | Age 70 - 74
M_AgeCat_11	Male | Age 75 - 79
M_AgeCat_12	Male | Age 80 - 84
M_AgeCat_13	Male | Age 85 - 89
M_AgeCat_14	Male | Age >=90
F_AgeCat_1	Female | Age 18 - 29
F_AgeCat_2	Female | Age 30 - 34
F_AgeCat_3	Female | Age 35 - 39
F_AgeCat_4	Female | Age 40 - 44
F_AgeCat_5	Female | Age 45 - 49
F_AgeCat_6	Female | Age 50 - 54
F_AgeCat_7	Female | Age 55 - 59
F_AgeCat_8	Female | Age 60 - 64
F_AgeCat_9	Female | Age 65 - 69
F_AgeCat_10	Female | Age 70 - 74
F_AgeCat_11	Female | Age 75 - 79
F_AgeCat_12	Female | Age 80 - 84
F_AgeCat_13	Female | Age 85 - 89
F_AgeCat_14	Female | Age >=90
Origin:	
TRNSFER	Transfer from another facility

Comorbidities:	
AIDS	Acquired immune deficiency syndrome
ALCOHOL	Alcohol abuse
ANEMDEF	Deficiency Anemias
CHF	Congestive heart failure
COAG	Coagulopathy
DM	Diabetes w/o chronic complications
DMCX	Diabetes w/ chronic complications
DRUG	Drug abuse
IMMUNE	Immune disorders
LIVER	Liver disease
LYTES	Fluid and electrolyte disorders
Major Diagnostic Categories (MDC):	
MDC_1	MDC 1: Nervous System
MDC_3	MDC 3: Ear Nose Mouth And Throat
MDC_4	MDC 4: Respiratory System
MDC_5	MDC 5: Circulatory System
MDC_6	MDC 6: Digestive System
MDC_7	MDC 7: Hepatobiliary System And Pancreas
MDC_8	MDC 8: Musculoskeletal And Connective
MDC_9	MDC 9: Skin Subcutaneous And Breast
MDC_10	MDC 10: Endocrine Nutritional And Metabolic
MDC_11	MDC 11: Kidney And Urinary Tract
MDC_13	MDC 13: Female Reproductive System
MDC_16	MDC 16: Blood and Immunological
MDC_17	MDC 17: Myeloproliferative Diseases and Disorders
MDC_18	MDC 18: Infectious and Parasitic
MDC_21	MDC 21: Injuries Poison And Toxic
METS	Metastatic cancer
PERIVASC	Peripheral vascular disease
PULMCIRC	Pulmonary circulation disease
RENLFAIL	Renal failure
VALVE	Valvular disease
WGHTLOSS	Weight loss
Modified Diagnostic Related Groups (MDRG):	
mdrg_1001	Adrenal & pituitary procedures
mdrg_1002	Amputation of lower limb for endocrine
mdrg_1003	O.R. procedures for obesity
mdrg_1004	Skin grafts & wound debridement for endocrine, nutrit, metab disorders
mdrg_1005	Thyroid parathyroid & thyroglossal procedures
mdrg_1006	Other endocrine nutrit & metab proc
mdrg_102	Craniotomy w major dev impl/acute complex CNS
mdrg_103	Craniotomy & endovascular intracranial procedures
mdrg_104	Spinal procedures
mdrg_105	Ventricular shunt procedures
mdrg_106	Carotid artery stent procedure
mdrg_107	Extracranial procedures
mdrg_108	Periph & cranial nerve & other nerv syst proc
mdrg_1101	Kidney transplant
mdrg_1102	Major bladder procedures
mdrg_1103	Kidney & ureter procedures for neoplasm
mdrg_1104	Kidney & ureter procedures for non-neoplasm
mdrg_1105	Minor bladder procedures
mdrg_1106	Prostatectomy
mdrg_1107	Transurethral procedures

mdrg_1108	Urethral procedures
mdrg_1109	Other kidney & urinary tract procedures
mdrg_1203	Testes procedures
mdrg_1204	Transurethral prostatectomy
mdrg_1301	Pelvic evisceration - rad hysterectomy
mdrg_1302	Uterine & adnexa proc ovarian or adnexal malig
mdrg_1303	Uterine adnexa proc non-ovarian/adnexal malig
mdrg_1304	Uterine & adnexa proc for non-malignancy
mdrg_1305	DnC conization laparoscopy & tubal interruption
mdrg_1306	Vagina cervix & vulva procedures
mdrg_1307	Female reproductive system reconstructive
mdrg_1308	Other female reproductive system procedures
mdrg_1601	Splenectomy
mdrg_1602	Other O.R. proc of the blood & blood forming
mdrg_1707	Lymphoma & leukemia
mdrg_1708	Lymphoma & non-acute leukemia
mdrg_1709	Myeloprolif disord or poorly diff neopl w maj OR proc
mdrg_1710	Myeloprolif disord or poorly diff neopl w other OR proc
mdrg_1801	Infectious & parasitic diseases w procedure
mdrg_1802	Postoperative or post-traumatic infections
mdrg_2101	Wound debridement for injuries
mdrg_2102	Skin grafts for injuries
mdrg_2103	Hand procedures for injuries
mdrg_2104	Other O.R. procedures for injuries
mdrg_2408	Other O.R. procedures for multiple sig trauma
mdrg_301	Acute major eye infections
mdrg_302	Other ear nose mouth & throat O.R. procedures
mdrg_303	Sinus & mastoid procedures
mdrg_304	Mouth procedures
mdrg_305	Salivary gland procedures
mdrg_401	Major chest procedures
mdrg_402	Other resp system O.R. procedures
mdrg_502	Perc cardiovasc proc w non-drug-eluting stent
mdrg_503	Cardiac valve & oth major cardiothoracic proc
mdrg_504	Cardiac defibrillator implant
mdrg_505	Other cardiothoracic procedures
mdrg_506	Coronary bypass w PTCA
mdrg_507	Coronary bypass w cardiac cath
mdrg_509	Amputation for circ sys disorders
mdrg_510	Permanent cardiac pacemaker implant
mdrg_511	Perc cardiovasc proc w drug-eluting stent
mdrg_513	Perc cardiovasc proc w/o coronary artery stent
mdrg_514	Other vascular procedures
mdrg_515	Upper limb & toe amputation
mdrg_516	Cardiac pacemaker device replacement
mdrg_517	Cardiac pacemaker revision
mdrg_519	Other circulatory system O.R. procedures
mdrg_601	Stomach esophageal & duodenal
mdrg_602	Major small & large bowel proc
mdrg_603	Rectal resection
mdrg_604	Peritoneal adhesiolysis
mdrg_605	Appendectomy w complicated principal diag
mdrg_606	Appendectomy w/o complicated principal diag
mdrg_607	Minor small & large bowel procedures
mdrg_608	Anal & stomal procedures

mdrg_609	Inguinal & femoral hernia procedures
mdrg_610	Hernia procedures except inguinal & femoral
mdrg_611	Other digestive system O.R. procedures
mdrg_701	Pancreas liver & shunt procedures
mdrg_702	Biliary tract proc except only cholecyst
mdrg_703	Cholecystectomy w c.d.e.
mdrg_704	Cholecystectomy except by laparoscope
mdrg_705	Laparoscopic cholecystectomy
mdrg_706	Hepatobiliary diagnostic procedures
mdrg_707	Other hepatobiliary or pancreas procedures
mdrg_7701	Heart transplant or implant heart assist sys
mdrg_801	Combined anterior/posterior spinal fusion
mdrg_802	Spinal fus exc cerv w spinal curv/malig/infec
mdrg_803	Spinal fusion except cervical
mdrg_804	Bilateral or multiple major joint procs
mdrg_805	Wnd debrid & skn grft exc hand for musculo
mdrg_806	Revision of hip or knee replacement
mdrg_807	Major joint replacement or reattachment
mdrg_808	Cervical spinal fusion
mdrg_809	Amputation for musculoskeletal sys
mdrg_810	Biopsies of musculoskeletal system
mdrg_811	Hip & femur procedures except major joint
mdrg_812	Major joint & limb reattachment
mdrg_813	Knee procedures w pdx of infection
mdrg_814	Knee procedures w/o pdx of infection
mdrg_815	Back & neck proc exc spinal fusion
mdrg_816	Lower extrem & humer proc
mdrg_817	Local excision & removal int fix devices
mdrg_819	Soft tissue procedures
mdrg_820	Foot procedures
mdrg_826	Other musculoskelet sys & conn tiss proc
mdrg_8899	Non-Extensive O.R. Proc Unrelated to PDX
mdrg_901	Skin graft &/or debrid for skn ulcer or cellulitis
mdrg_902	Skin graft &/or debrid exc for skin ulcer
mdrg_903	Other skin subcut tiss & breast

c-statistic = .769

Source: http://qualityindicators.ahrq.gov/Modules/psi_resources.aspx
Parameter estimates are also included with the Technical Specifications attached in section S.2b

S.15. Detailed risk model specifications *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Available in attached Excel or csv file at S.2b

S.15a. Detailed risk model specifications (*if not provided in excel or csv file at S.2b*)
Available in attached Excel file at S.2b

S.16. Type of score:
Rate/proportion
If other:

S.17. 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.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
The observed rate is the number of discharge records where the patient experienced the PSI adverse event divided by the number of discharge records at risk for the event. The expected rate is a comparative rate that incorporates information about a reference population that is not part of the user's input dataset – what rate would be observed if the expected level of care observed in the reference population and estimated with risk adjustment regression models, were applied to the mix of patients with demographic and comorbidity distributions observed in the user's dataset. The expected rate is calculated only for risk-adjusted indicators.

The expected rate is estimated for each person using a generalized estimating equations (GEE) approach to account for correlation at the hospital or provider level.

The risk-adjusted rate is a comparative rate that also incorporates information about a reference population that is not part of the input dataset – what rate would be observed if the level of care observed in the user's dataset were applied to a mix of patients with demographics and comorbidities distributed like the reference population? The risk adjusted rate is calculated using the indirect method as observed rate divided by expected rate multiplied by the reference population rate. The smoothed rate is the weighted average of the risk-adjusted rate from the user's input dataset and the rate observed in the reference population; the smoothed rate is calculated with a shrinkage estimator to result in a rate near that from the user's dataset if the provider's rate is estimated in a stable fashion with minimal noise, or to result in a rate near that of the reference population if the variance of the estimated rate from the input dataset is large compared with the hospital-to-hospital variance estimated from the reference population. Thus, the smoothed rate is a weighted average of the risk-adjusted rate and the reference population rate, where the weight is the signal-to-noise ratio. In practice, the smoothed rate brings rates toward the mean, and tends to do this more so for outliers (such as rural hospitals).

For additional information, please see the supplemental materials for the AHRQ QI Empirical Methods.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Not applicable

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not applicable

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Exclude cases with missing gender (SEX=missing), age (AGE=missing), quarter (DQTR=missing), year (YEAR=missing), or principal diagnosis (DX1=missing). Missingness on these variables, in aggregate, almost never exceeds 1% of eligible records.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
While the measure is tested and specified using data from the Healthcare Cost and Utilization Project (HCUP) (see section 1.1 and 1.2 of the measure testing form), the measure specifications and software are specified to be used with any ICD-9-CM- or ICD-10-

CM/PCS coded administrative billing/claims/discharge dataset.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
Not applicable

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
PSI09_Testing_Form_160513_v03.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 2909
Measure Title: Perioperative Hemorrhage or Hematoma Rate (PSI 09)
Date of Submission: 5/13/2016
Type of Measure:

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

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in

understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]
AND
If patient preference (e.g., informed decision making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):
- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR
- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;
OR
there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes
10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).
11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.
12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of

exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.
14. Risk factors that influence outcomes should not be specified as exclusions
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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.23*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ administrative claims	☒ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

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

All analyses were completed using data from the Healthcare Cost and Utilization Project (HCUP) State Inpatient Databases (SID), 2011-2013. HCUP is a family of health care databases and related software tools and products developed through a Federal-State-Industry partnership and sponsored by the Agency for Healthcare Research and Quality (AHRQ).[1] HCUP databases bring together the data collection efforts of State data organizations, hospital associations, private data organizations, and the Federal government to create a national information resource of encounter-level health care data. The HCUP SID contain the universe of the inpatient discharge abstracts in participating States, translated into a uniform format to facilitate multi-State comparisons and analyses. All states provide data for community hospitals and together, the SID encompasses about 97 percent of all U.S. community hospital discharges. For the analyses presented here, we use 34 states representing about 89 percent of the U.S. community hospital discharges, for a total of about 30 million hospital discharges from community hospitals. As defined by the American Hospital Association, community hospitals are all non-Federal, short-term, general or other specialty hospitals, excluding hospital units of institutions. Included among community hospitals are public and academic medical centers, specialty hospitals such as obstetrics–gynecology, ear–nose–throat, orthopedic and pediatric institutions. Short-stay

[1] HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2011-2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

rehabilitation, long-term acute care hospitals are excluded from the data used for the reported analyses.

Each of the 34 states included in the dataset report information about whether a diagnosis was present on admission (POA) and information on the timing of procedures during the hospitalization. POA data[2] is important to distinguish complications that occur in-hospital from diagnoses that existed prior to hospitalization. Edit checks on POA were developed using a separate analysis of HCUP databases that examined POA coding in the 2013 SID at hospitals that were required to report POA to CMS. The edits identify general patterns of suspect reporting of POA. The edits do not evaluate whether a valid POA value (e.g., Y or N) is appropriate for the specific diagnosis. There are three hospital-level edit checks:

1. Indication that a hospital has POA reported as Y on all diagnoses on all discharges
2. Indication that a hospital has POA reported as missing on all non-Medicare discharges
3. Indication that a hospital reported POA as missing on all nonexempt diagnoses for 15 percent or more of discharges. The cut-point of 15 percent was determined by 2 times the standard deviation plus the mean of the percentage for hospitals required to report POA to CMS.

Hospitals that failed any of the edit checks were excluded from the dataset.
The SID data elements include International Classification of Diseases, 9th Revision, Clinical Modification coded principal and secondary diagnoses and procedures, additional detailed clinical and service information based on revenue codes, admission and discharge status, patient demographics, expected payment source (Medicare, Medicaid, private insurance as well as the uninsured), total charges and length of stay (www.hcup-us.ahrq.gov).

1.3. What are the dates of the data used in testing?
HCUP data included 2011-2013, for most tests we combine SID data for 2 years prior to calculating rates and testing the measure. This is termed "2-year pooled data" in the results below.

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

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.26*)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

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

[2] Present-on –Admission (POA) was added as a data element to the uniform bill form (UB-04) effective October 1, 2007, and hospitals incurred a payment penalty for not including POA on Medicare records beginning October 1, 2008. Each of the several diagnoses in a discharge record can be flagged as "present at the time the order for inpatient admission occurs" or not (see http://www.cdc.gov/nchs/icd/icd9cm_addenda_guidelines.htm).

Table 1. Reference Population Rate and Distribution of Hospital Performance for PSI 09 Perioperative Hemorrhage or Hematoma Rate in 2-year Pooled Data (2011-2013)

Overall Reference Population Rate				
Year[3]	Number of Hospitals	Outcome of Interest (Numerator)[1]	Population at Risk (Denominator)[1]	Observed Rate Per 1000 Surgical Discharges[1]
2011-2012	3,432	52,548	11,043,434	4.7583
2012-2013	3,613	48,663	10,780,407	4.5140

Distribution of Hospital-level Observed Rates in Reference Population								
Year[3]	Number of Hospitals	Rates per 1000 Surgical Discharges (p=percentile)[2]						
		Mean	SD[2]	p5	p25	Median	p75	p95
2011-2012	3,432	3.81	3.83	0.00	1.56	3.52	5.21	8.90
2012-2013	3,613	3.64	4.13	0.00	1.32	3.28	5.05	8.50

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2011 - 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov (AHRQ QI Software Version 6.0)

[1]The observed rate refers to the total rate for all observations included in the reference population data (numerator) divided by the total combined eligible population of all hospitals included in the reference population data (denominator).

[2]The distribution of hospital rates reports the mean and standard deviation (SD) of the observed rates for all hospitals included in the dataset with at least one case in the denominator, as well as the observed rate for hospitals in the 5th, 25th, 50th (median), 75th, and 95th percentile. Standard deviation refers to the spread in observed values in relation to the mean.

[3] Reference population is limited to states with present on admission data (POA). Since many states did not report POA data prior to 2011 we have not included testing prior to 2011.

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

See 1.5 (Table 1)

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.

Some tests require comparisons across two or three years of data (2011-2013). When no comparisons are required for the test, typically 2013 data are used. Some validity testing uses only 2011 or 2012 data.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Age and sex were the only patient-level sociodemographic variables that were available and analyzed in the data used for measure development and testing. Many of the HCUP SID include race/ethnicity, and all of the HCUP SID include the primary expected source of payment, and zip code of residence, which could be used to capture socioeconomic characteristics at an ecological (community) level. While some of these variables were used to assess disparities at the national level, these variables were not used in the current risk adjustment model, based on our conceptual description (i.e., logical rationale or theory informed by literature and content experts) of the causal pathway between these factors, patient clinical factors, quality of care, and outcome, described in Section 2b4.3 below.

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*)

Signal-to-Noise. The signal-to-noise ratio is a measure of reliability that is calculated at the hospital level and then summarized across the entire population of US hospitals. It compares the degree to which risk adjusted rates are different from hospital to hospital (the signal) to how precise the rates are within hospitals (the noise). This metric is a stringent measure of reliability that takes into account the observed distribution of risk adjusted rates within a reference population. An indicator with a low signal-to-noise ratio may not be able to distinguish differences in performance among hospitals, or may identify differences inconsistently within the same time period. An indicator with a high signal-to-noise ratio will be more likely to consistently distinguish performance differences among hospitals (e.g. one hospital performs better than others).

The signal-to-noise ratio is estimated for each hospital. The overall signal-to-noise estimate is an average of hospital-level signal-to-noise ratios weighted by a value of one divided by the signal plus the hospital's noise for PSI 09. Hospitals with smaller denominators (the number of patients at risk) will have lower weight, and less influence on the overall signal-to-noise ratio, because of higher noise. Weighting reduces the influence of hospitals that have less reliable rates due to very small denominators (the number of patients at risk) on the overall signal-to-noise ratio estimate.

Because the signal-to-noise ratio quantifies the ability to consistently discriminate one hospital's performance from the other hospitals in the population, it is sensitive to the distribution of hospital sizes as well as the distribution of risk-adjusted rates in the reference population. If the hospitals in a population all have performance in a narrow range (low signal), it is more difficult to reliably distinguish among hospitals' performance than when hospital performance is spread out over a much wider range (high signal). For example, if all hospitals have nearly perfect performance, it will be impossible to distinguish among them. As a consequence, if the distribution of hospital rates changes over time, the signal-to-noise ratio will also change.

There is no universally accepted threshold of "adequate" signal-to-noise ratio. Different methods of calculating reliability and signal-to-noise (e.g., split sample or test-retest reliability of the data, different methods of calculating the hospital signal-to-noise ratio) result in different distributions of reliability scores. In addition, "adequate" depends on the specific application and judgment of the user. For instance, if a complication such as mortality is very important (e.g. leads to great harm to the patient) a lower reliability may be acceptable. However, the AHRQ QI program generally considers ratios between 0.4 – 0.8 as acceptable. It is rare to achieve reliability above 0.8, using hospital signal-to-noise ratios as an indicator of reliability. To account for the uncertainty (noise) in a hospital's performance due to low volume, a longer period of data can be used or smoothed rates can be calculated.

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)

Table 2. Signal-to-Noise Ratio by Hospital Size Decile, PSI 09 Perioperative Hemorrhage or Hematoma Rate Using 2-year pooled data (2012-2013)

Hospital Size Decile	Number of Hospitals	Avg. Number of Discharges per Hospital in Decile	Avg. Signal-to-Noise Ratio for Hospitals in Decile
1 (smallest)	357	43.9	0.0342
2	357	219.7	0.0893
3	358	521.4	0.1822
4	357	957.2	0.2741
5	357	1,533.5	0.3796
6	358	2,257.4	0.4778
7	357	3,131.7	0.5630
8	358	4,308.4	0.6456
9	357	6,202.5	0.7349
10 (largest)	357	11,001.4	0.8399
Overall	3,573	3,017.2	0.6661

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2012 - 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov (AHRQ QI Software Version 6.0)

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?*)
Signal-to-noise ratios were smaller for hospitals with fewer than 1,534 qualifying discharges per year (average signal-to-noise ratio less than 0.38). Smoothed rates, which are recommended for all hospitals (and are implemented in the AHRQ software), address reliability concerns particularly for small hospitals. Hospitals with more than 2257 qualifying discharges on average have risk adjusted rates with moderate to high reliability (average signal-to-noise ratio of 0.48 to 0.84). Overall, the signal-to-noise ratio for this indicator is good with an overall signal-to-noise ratio of 0.67.

2b2. VALIDITY TESTING

2b2.1. 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*)

2b2.2. 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)*

Critical Data Elements

Much of the validity assessments of PSI 09, particularly of the validity of the PSI 09 algorithm and administrative data to capture postoperative hemorrhage or hematoma, have been published in the peer review

literature. We summarize the most relevant literature in this form, and provide a full evidence summary in the attached Evidence Form.

Performance Measure Score

Systematic Assessment of Face Validity

We utilized a structured panel review to evaluate face validity (from a clinical perspective) of the Patient Safety Indicators. The panels were convened in 2001 and 2002. It is anticipated that the results of face validity review would be similar if panels were convened in more recent years, given that the clinical characteristics of these events, treatment and prevention approaches, and sequelae have not changed substantially since 2002. The clinical panel review process was based on the RAND appropriateness method, a modified Delphi process also known as a nominal group technique.

Twenty-one professional clinical organizations were invited to submit nominations. These organizations were selected based on the applicability of the specialty or subspecialty to potential Patient Safety Indicators. Clinical areas represented by the panels included internal medicine, cardiology, radiology, geriatrics, surgical and critical care nursing, anesthesiology, pharmacy, inpatient medicine and surgery (including thoracic, neurology, orthopedic, colorectal, urology, spine, and transplant surgical subspecialties). For assignments to each panel, a list of applicable specialties was identified for the indicators to be evaluated by that panel. Panelists were selected so that each panel had diverse membership in terms of practice characteristics and setting. For PSI 09, 7 members of a multispecialty panel and 6 members of a surgical subspecialty panel completed the evaluation in full. Additional details of panel composition are available online at http://archive.ahrq.gov/clinic/tp/hospdatp.htm.

Panelists completed a 10-item questionnaire, tailored to each specific indicator. Following the initial rating of the indicators, panelists participated in a moderated 90-minute conference call, where opinions about the indicators were discussed. The panelists then completed the same 10-item questionnaire again, and submitted their final ratings. Ratings were summarized in accordance with the RAND Appropriateness Method.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Critical Data Elements

Early studies of the validity of the PSI 09 algorithm for identifying postoperative hemorrhage and hematoma found moderate to high positive predictive value (PPV) of 75% (95% CI 66% to 83%) **in a sample of VA hospitals**[3] **and 95% (95% CI, 90–98%) in 35 volunteer nonfederal hospitals.**[4] **However, in the latter study the PPV from the clinical perspective (reflecting a hemorrhage or hematoma clinically related to a procedure, requiring a second procedure to treat the hemorrhage or hematoma) was 78% (95% CI, 58–90%).**[4] These studies were performed prior to the implementation of POA flags; a studying utilizing POA found a PPV of 97% (95% CI 85% to 100%) in a single site[5].

After evaluating the early evidence relating to PSI 09, AHRQ expanded the list of procedure codes that qualify as treatment for hemorrhage or hematoma (v4.4). In the sample of volunteer hospitals, this updated specification improved the indicator's estimated sensitivity from 42% (95% CI, 23-64%) to 85% (95% CI, 67-94%), while maintaining the PPV at 76% (95% CI, 60-88%).[4]

Systematic Assessment of Face Validity

The multi-specialty Panel and Surgical Panel both rated the indicator as acceptable on overall usefulness as an indicator of potentially preventable complications of care.

Table 3. Clinician Panel Evaluations of the Face Validity of PSI 09[4]

Multi-specialty Panel (MSP) Evaluation			Surgical Panel (SP) Evaluation		
Overall Rating[1]	Agreement[2]	Acceptability[3]	Overall Rating[1]	Agreement[2]	Acceptability[3]
7	Indeterminate	Acceptable	7	Agreement	Acceptable **(-)**

[1]Median panel overall rating of the indicator on a scale from 1 to 9, with the higher rating indicating better measurement
[2]Level of agreement, where "agreement" corresponds to little dispersion of opinion, "indeterminate" means that the opinion ranged but did not reach the point of clear "disagreement", the final category where there were panelists with diametrically different opinions
[3]"Acceptable" indicates that the indicator was rated as useful by almost all panelists. "Acceptable (-)" indicates that the indicator was rated as useful by most panelists, although a few rated it as less useful (but not as poor). "Unclear" indicates that panelists rated the usefulness of the indicator as moderate. For further details of methods, see http://archive.ahrq.gov/clinic/tp/hospdatp.htm
[4]PSI 09 was evaluated under a previous name (i.e. Postoperative Hemorrhage or Hematoma Rate).

2b2.4. 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?*)

Studies in the peer-review literature reported moderate to high PPV from 75% to 95%, depending on the clinical setting. The sensitivity of this measure was improved to 42% after expansion of the procedure codes that qualify as treatment for hemorrhage or hematoma, without reducing specificity.

In our clinical panel review, the indicator had high face validity for use in quality improvement and hospital comparative assessments.

2b3. EXCLUSIONS ANALYSIS

[3] Borzecki AM, Kaafarani H, Cevasco M, et al. How valid is the AHRQ Patient Safety Indicator "postoperative hemorrhage or hematoma"? *J Am Coll Surg.* 2011;212(6):946-953 e941-942.
[4] Utter GH, Zrelak PA, Baron R, et al. Detecting postoperative hemorrhage or hematoma from administrative data: the performance of the AHRQ Patient Safety Indicator. *Surgery.* 2013;154(5):1117-1125.
[5] Ramanathan R, Leavell P, Stockslager G, Mays C, Harvey D, Duane TM. Validity of Agency for Healthcare Research and Quality Patient Safety Indicators at an academic medical center. *Am Surg.* 2013;79(6):578-582.

NA ☐ no exclusions — *skip to section 2b4*

2b3.1. 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*)
Empirical Evaluation of Exclusions: Using the 2013 data from 34 states, we examined the percent of potential denominator cases excluded by each criterion as listed in the measure specifications.

2b3.2. 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 4 shows the results of the most recent exclusions analysis.

Table 4. Number and Percent of Discharges Excluded, by Denominator Exclusion Criteria, PSI 09 Perioperative Hemorrhage or Hematoma Rate[1]

PSI 09	Denominator			Potential Numerator[2]		
Exclusion Name	Exclusion Count	After Exclusions	% Change	Exclusion Count	After Exclusions	% Change
No Exclusions applied	-	**6,671,854**	-	-	**38,704**	-
Exclude Principal Diagnosis of Perioperative Hemorrhage or Hematoma	12,223	6,659,631	0.2%	488	38,216	1.3%
Exclude if control of perioperative hemorrhage or Miscellaneous hemorrhage hematoma-related procedure are the only OR procedures	175,844	6,496,010	2.6%	4,080	34,624	10.5%
Exclude if control of perioperative hemorrhage or Miscellaneous Hemorrhage or hematoma-related procedure occurs before the first OR procedure;	5,836	6,666,018	0.1%	5,836	32,868	15.1%
Exclude MDC 14	1,050,160	5,621,694	15.7%	327	38,377	0.8%
Exclude Coagulation Disorders	254,874	6,416,980	3.8%	6,720	31,984	17.4%

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov (AHRQ QI Software Version 6.0)
[1]This indicator does not have numerator exclusion criteria.
[2]Potential numerator cases are those that would have qualified for the numerator if not for a particular denominator exclusion criterion.

2b3.3. 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)

As part of the inherent design of the indicator and to minimize false positive events, the numerator limits detection to hospitalizations involving a hemorrhage or hematoma diagnosis in the context of a putatively reparative procedure. The intent is to isolate those hemorrhages that can truly be linked to a surgical procedure and are of sufficient severity to be consequential to the patient.

The denominator excludes cases of hemorrhage or hematoma that were present on admission, when the only operating room procedure was for treatment of hemorrhage or hematoma, or when the procedure for treatment of hemorrhage or hematoma occurred before the first operating room procedure. In each of these cases, the precipitating events happened before the targeted hospitalization, and are thus unlikely to reflect quality of care during that hospitalization. This indicator also excludes hemorrhage and hematoma associated with pregnancy, childbirth and puerperium as it more difficult in this setting to differentiate between normal and abnormal bleeding that is under the control of the healthcare team. Patients with coagulation disorders are excluded as hemorrhage in these patients may be less preventable. The empirical analysis supports these exclusions, as they capture a non-trivial number of numerator events.

2b4. 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 2b5.

2b4.1. What method of controlling for differences in case mix is used?
☐ **No risk adjustment or stratification**
☒ **Statistical risk model with 23 risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.
Not applicable

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

Clinical Factors
For PSI 09, we considered a standard set of covariates grouped into four categories: demographics, severity of illness, comorbidities and transfer-in status. Covariates that were considered as potential risk adjusters included gender and age, MDC, Modified Diagnostic Related Groups (MDRGs) (defined as the base MS-DRG without comorbidity or complication distinctions), AHRQ Comorbidity Software categories and whether they were transferred from another facility. Only those covariates with at least 30 cases for PSI 09 are retained. A parsimonious model was identified using backward stepwise selection with bootstrapping.

The omitted covariate within mutually exclusive categories is the reference group for those categories. Reference categories are usually 1) the most common and/or 2) the least risk. The choice of omitted reference category does not affect predicted probabilities or model performance.

For the MDRGs, the risk reported is the residual risk after adjustment for the MDC to which the MDRG belongs. Likewise, the risk reported for MDCs represents the average risk of all MSDRGs in that MDC not included in the model.

Additional details are available in the *AHRQ Quality Indicator Empirical Methods* document, included in the supplemental file and available on the AHRQ QI website.

Sociodemographic Factors

Empirical studies have generally demonstrated minimal differences in PSI 09 rates across racial/ethnic categories (e.g., 2.00, 1.99, and 1.83 events per 1,000 white, African American, and Hispanic patients, respectively, in the Nationwide Inpatient Sample).[6] More importantly, there is no evidence or causal model to suggest that SDS factors are associated with perioperative hemorrhage or hematoma independent of quality of care, or are mediated by pre-hospital care (which may not fall within the proper realm of hospital accountability). Accordingly, consistent with the guidance provided by NQF in the SDS Trial Period FAQs, AHRQ believes that it would be inappropriate to include other SDS variables in the risk-adjustment approach for PSI 09, which is an in-hospital outcome measure.

2b4.4a. What were the statistical results of the analyses used to select risk factors?

The current risk adjustment coefficients for PSI 09 can be found attached to the technical specifications document. The risk model includes 170 risk categories, including 26 age-gender categories in 5-year age categories between ages 30 and 89, and 2 age-gender categories ranging from below age 30 (i.e. 18-29) as one category and ages 90+ as another category, transfer in from another acute care facility and 17 comorbidities. The remainder of selected risk factors account for the reason for admission and the type of surgery that was performed during the hospitalization, including MDC and MS-DRGs collapsed to remove Complication or Comorbidity/ Major Complication or Comorbidity (CC/MCC) distinctions.

Table 5. Risk Adjustment Coefficients for PSI 09 Perioperative Hemorrhage or Hematoma Rate

Parameter	Label	DF	Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square
Intercept	Intercept	1	-6.0223	0.0702	7350.088	<.0001
Sex | Age Demographics						
M_AgeCat_1	Male | Age 18 - 29	1	-0.1313	0.0745	3.1050	0.0781
M_AgeCat_2	Male | Age 30 - 34	1	0.0611	0.0809	0.5703	0.4502
M_AgeCat_3	Male | Age 35 - 39	1	-0.0535	0.0720	0.5532	0.4570
M_AgeCat_4	Male | Age 40 - 44	1	-0.0190	0.0593	0.1027	0.7486
M_AgeCat_5	Male | Age 45 - 49	1	-0.0197	0.0489	0.1618	0.6875
M_AgeCat_6	Male | Age 50 - 54	1	0.0262	0.0414	0.3984	0.5279
M_AgeCat_7	Male | Age 55 - 59	1	-0.0189	0.0383	0.2433	0.6218
M_AgeCat_8	Male | Age 60 - 64	1	0.1078	0.0394	7.4950	0.0062
M_AgeCat_9	Male | Age 65 - 69	1	0.1275	0.0386	10.9276	0.0009
M_AgeCat_10	Male | Age 70 - 74	1	0.0364	0.0370	0.9696	0.3248
M_AgeCat_11	Male | Age 75 - 79	1	0.0806	0.0397	4.1157	0.0425
M_AgeCat_12	Male | Age 80 - 84	1	0.0045	0.0453	0.0101	0.9200
M_AgeCat_13	Male | Age 85 - 89	1	-0.0913	0.0612	2.2282	0.1355
M_AgeCat_14	Male | Age >=90	1	-0.1980	0.1076	3.3860	0.0658
F_AgeCat_1	Female | Age 18 - 29	1	-0.2720	0.0693	15.4019	<.0001
F_AgeCat_2	Female | Age 30 - 34	1	0.0090	0.0635	0.0202	0.8869

[6] Romano, P. S., Geppert, J. J., Davies, S., Miller, M. R., Elixhauser, A., & McDonald, K. M. (2003). A national profile of patient safety in US hospitals. *Health Affairs*, *22*(2), 154-166.

Parameter	Label	DF	Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square
F_AgeCat_3	Female | Age 35 - 39	1	0.0424	0.0545	0.6071	0.4359
F_AgeCat_4	Female | Age 40 - 44	1	0.0078	0.0484	0.0257	0.8726
F_AgeCat_5	Female | Age 45 - 49	1	-0.0439	0.0455	0.9289	0.3352
F_AgeCat_6	Female | Age 50 - 54	1	-0.0543	0.0435	1.5613	0.2115
F_AgeCat_7	Female | Age 55 - 59	1	0.0037	0.0425	0.0077	0.9300
F_AgeCat_8	Female | Age 60 - 64	1	-0.0166	0.0413	0.1618	0.6875
F_AgeCat_9	Female | Age 65 - 69	*Reference Group				
F_AgeCat_10	Female | Age 70 - 74	1	0.0178	0.0419	0.1812	0.6703
F_AgeCat_11	Female | Age 75 - 79	1	0.1388	0.0439	10.0123	0.0016
F_AgeCat_12	Female | Age 80 - 84	1	-0.0782	0.0489	2.5648	0.1093
F_AgeCat_13	Female | Age 85 - 89	1	0.0178	0.0602	0.0880	0.7668
F_AgeCat_14	Female | Age >=90	1	-0.3619	0.0966	14.0293	0.0002
Origin						
TRNSFER	Transfer from another facility	1	0.1323	0.0285	21.5140	<.0001
Comorbidities						
AIDS	Acquired immune deficiency syndrome	1	-0.7499	0.2317	10.4753	0.0012
ALCOHOL	Alcohol abuse	1	0.1247	0.0385	10.4646	0.0012
ANEMDEF	Deficiency Anemias	1	-0.3240	0.0229	200.3910	<.0001
CHF	Congestive heart failure	1	0.2039	0.0308	43.8859	<.0001
COAG	Coagulopathy	1	0.3167	0.0823	14.8002	0.0001
DM	Diabetes w/o chronic complications	1	-0.1171	0.0179	42.6889	<.0001
DMCX	Diabetes w/ chronic complications	1	-0.1632	0.0322	25.6385	<.0001
DRUG	Drug abuse	1	0.1363	0.0457	8.8924	0.0029
IMMUNE	Immune disorders	1	0.4095	0.0275	221.0441	<.0001
LIVER	Liver disease	1	0.2788	0.0381	53.4808	<.0001
LYTES	Fluid and electrolyte disorders	1	-0.0931	0.0236	15.5649	<.0001
METS	Metastatic cancer	1	0.0869	0.0358	5.8851	0.0153
PERIVASC	Peripheral vascular disease	1	0.1953	0.0228	73.6165	<.0001
PULMCIRC	Pulmonary circulation disease	1	0.3020	0.0477	40.0222	<.0001
RENLFAIL	Renal failure	1	0.1449	0.0234	38.4997	<.0001
VALVE	Valvular disease	1	0.3796	0.0341	123.8486	<.0001
WGHTLOSS	Weight loss	1	0.2420	0.0311	60.4179	<.0001
Major Diagnostic Categories (MDC)						
MDC_1	MDC 1: Nervous System	1	1.5839	0.0971	266.0482	<.0001
MDC_3	MDC 3: Ear Nose Mouth And Throat	1	2.2971	0.0958	575.2025	<.0001

Parameter	Label	DF	Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square
MDC_4	MDC 4: Respiratory System	1	2.1134	0.0988	457.6746	<.0001
MDC_5	MDC 5: Circulatory System	1	2.9018	0.0827	1230.5107	<.0001
MDC_6	MDC 6: Digestive System	1	2.4435	0.1106	487.6702	<.0001
MDC_7	MDC 7: Hepatobiliary System And Pancreas	1	2.8237	0.1044	731.1700	<.0001
MDC_8	MDC 8: Musculoskeletal And Connective	1	-0.0868	0.1167	0.5530	0.4571
MDC_9	MDC 9: Skin Subcutaneous And Breast	1	1.8089	0.0800	511.9115	<.0001
MDC_10	MDC 10: Endocrine Nutritional And Metabolic	1	2.8813	0.1368	443.5266	<.0001
MDC_11	MDC 11: Kidney And Urinary Tract	1	2.5169	0.1461	296.8100	<.0001
MDC_13	MDC 13: Female Reproductive System	1	2.6529	0.3900	46.2686	<.0001
MDC_16	MDC 16: Blood and Immunological	1	2.3626	0.2386	98.0589	<.0001
MDC_17	MDC 17: Myeloproliferative Diseases and Disorders	1	2.5655	0.2500	105.2988	<.0001
MDC_18	MDC 18: Infectious and Parasitic	1	1.6559	0.1392	141.4778	<.0001
MDC_21	MDC 21: Injuries Poison And Toxic	1	2.0728	0.2303	81.0245	<.0001
Modified Diagnostic Related Groups (MDRG)						
mdrg_1001	Adrenal & pituitary procedures	1	-2.0639	0.1855	123.7639	<.0001
mdrg_1002	Amputation of lower limb for endocrine	1	-2.9065	0.1737	280.0458	<.0001
mdrg_1003	O.R. procedures for obesity	1	-2.2228	0.1297	293.4948	<.0001
mdrg_1004	Skin grafts & wound debridement for endocrine nutrit & metab disorders	1	-3.2915	0.2783	139.9150	<.0001
mdrg_1005	Thyroid parathyroid & thyroglossal procedures	1	-1.3892	0.1327	109.6033	<.0001
mdrg_1006	Other endocrine nutrit & metab proc	1	-3.1649	0.2104	226.3496	<.0001
mdrg_102	Craniotomy w major dev impl/acute complex CNS	1	-1.8913	0.2028	86.9714	<.0001
mdrg_103	Craniotomy & endovascular intracranial procedures	1	-1.4852	0.1053	198.7478	<.0001

Parameter	Label	DF	Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square
mdrg_104	Spinal procedures	1	-1.1351	0.1322	73.7534	<.0001
mdrg_105	Ventricular shunt procedures	1	-3.1970	0.4149	59.3700	<.0001
mdrg_106	Carotid artery stent procedure	1	-1.8035	0.2135	71.3259	<.0001
mdrg_107	Extracranial procedures	1	-0.2168	0.0825	6.8970	0.0086
mdrg_108	Periph & cranial nerve & other nerv syst proc	1	-1.6702	0.1636	104.2688	<.0001
mdrg_1101	Kidney transplant	1	-0.9051	0.1531	34.9381	<.0001
mdrg_1102	Major bladder procedures	1	-1.7668	0.1781	98.3696	<.0001
mdrg_1103	Kidney & ureter procedures for neoplasm	1	-1.5702	0.1459	115.8622	<.0001
mdrg_1104	Kidney & ureter procedures for non-neoplasm	1	-1.9346	0.1502	165.8917	<.0001
mdrg_1105	Minor bladder procedures	1	-0.6446	0.1725	13.9640	0.0002
mdrg_1106	Prostatectomy	1	-2.1176	0.2707	61.1888	<.0001
mdrg_1107	Transurethral procedures	1	-3.1277	0.1766	313.4877	<.0001
mdrg_1108	Urethral procedures	1	-2.0543	0.3432	35.8253	<.0001
mdrg_1109	Other kidney & urinary tract procedures	1	-1.9794	0.1589	155.2625	<.0001
mdrg_1203	Testes procedures	1	1.1874	0.2074	32.7756	<.0001
mdrg_1204	Transurethral prostatectomy	1	1.1372	0.1004	128.3885	<.0001
mdrg_1301	Pelvic evisceration - rad hysterectomy	1	-1.4814	0.4042	13.4307	0.0002
mdrg_1302	Uterine & adnexa proc ovarian or adnexal malig	1	-1.6315	0.4011	16.5458	<.0001
mdrg_1303	Uterine adnexa proc non-ovarian/adnexal malig	1	-1.7890	0.3958	20.4286	<.0001
mdrg_1304	Uterine & adnexa proc for non-malignancy	1	-1.4301	0.3857	13.7493	0.0002
mdrg_1305	DnC conization laparoscopy & tubal interruption	1	-2.2780	0.4302	28.0380	<.0001
mdrg_1306	Vagina cervix & vulva procedures	1	-1.6950	0.4055	17.4704	<.0001
mdrg_1307	Female reproductive system reconstructive	1	-2.1745	0.4105	28.0608	<.0001
mdrg_1308	Other female reproductive system procedures	1	-1.3589	0.4236	10.2931	0.0013
mdrg_1601	Splenectomy	1	-0.9241	0.2984	9.5938	0.0020
mdrg_1602	Other O.R. proc of the blood & blood forming	1	-1.4640	0.2959	24.4761	<.0001

Parameter	Label	DF	Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square
mdrg_1707	Lymphoma & leukemia	1	-1.1088	0.2608	18.0735	<.0001
mdrg_1708	Lymphoma & non-acute leukemia	1	-2.3547	0.2980	62.4376	<.0001
mdrg_1709	Myeloprolif disord or poorly diff neopl w maj OR proc	1	-0.9244	0.2643	12.2287	0.0005
mdrg_1710	Myeloprolif disord or poorly diff neopl w other OR proc	1	-2.2141	0.3770	34.4918	<.0001
mdrg_1801	Infectious & parasitic diseases w procedure	1	-1.1477	0.1312	76.4868	<.0001
mdrg_1802	Postoperative or post-traumatic infections	1	-1.1583	0.1508	58.9707	<.0001
mdrg_2101	Wound debridements for injuries	1	-1.6348	0.2913	31.5022	<.0001
mdrg_2102	Skin grafts for injuries	1	-1.0491	0.2662	15.5291	<.0001
mdrg_2103	Hand procedures for injuries	1	-2.7242	0.4647	34.3676	<.0001
mdrg_2104	Other O.R. procedures for injuries	1	-1.2233	0.2319	27.8309	<.0001
mdrg_2408	Other O.R. procedures for multiple sig tr	1	0.5487	0.1277	18.4609	<.0001
mdrg_301	Acute major eye infections	1	-1.5232	0.1160	172.4425	<.0001
mdrg_302	Other ear nose mouth & throat O.R. procedures	1	-1.1060	0.1140	94.2124	<.0001
mdrg_303	Sinus & mastoid procedures	1	-1.9228	0.4154	21.4295	<.0001
mdrg_304	Mouth procedures	1	-2.0467	0.3415	35.9092	<.0001
mdrg_305	Salivary gland procedures	1	-1.5915	0.3108	26.2228	<.0001
mdrg_401	Major chest procedures	1	-1.1686	0.0870	180.4056	<.0001
mdrg_402	Other resp system O.R. procedures	1	-2.8403	0.1468	374.5334	<.0001
mdrg_502	Perc cardiovasc proc w non-drug-eluting stent	1	-3.2921	0.1105	887.5255	<.0001
mdrg_503	Cardiac valve & oth maj cardiothoracic proc	1	-0.7394	0.0584	160.2279	<.0001
mdrg_504	Cardiac defibrillator implant	1	-2.8727	0.1046	754.1971	<.0001
mdrg_505	Other cardiothoracic procedures	1	-0.4937	0.0883	31.2668	<.0001
mdrg_506	Coronary bypass w PTCA	1	-1.4684	0.1608	83.4246	<.0001
mdrg_507	Coronary bypass w cardiac cath	1	-1.4800	0.0616	576.4492	<.0001
mdrg_509	Amputation for circ sys disorders	1	-2.3993	0.1016	557.7818	<.0001

Parameter	Label	DF	Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square
mdrg_510	Permanent cardiac pacemaker implant	1	-2.7936	0.1401	397.8469	<.0001
mdrg_511	Perc cardiovasc proc w drug-eluting stent	1	-3.5340	0.0740	2278.1001	<.0001
mdrg_513	Perc cardiovasc proc w/o coronary artery stent	1	-2.9842	0.0972	943.5254	<.0001
mdrg_514	Other vascular procedures	1	-1.6339	0.0615	705.4056	<.0001
mdrg_515	Upper limb & toe amputation	1	-4.4025	0.4119	114.2164	<.0001
mdrg_516	Cardiac pacemaker device replacement	1	-3.4688	0.4509	59.1717	<.0001
mdrg_517	Cardiac pacemaker revision	1	-2.4123	0.1712	198.6594	<.0001
mdrg_519	Other circulatory system O.R. procedures	1	-2.7269	0.1167	546.2852	<.0001
mdrg_601	Stomach esophageal & duodenal	1	-1.6443	0.1046	247.2597	<.0001
mdrg_602	Major small & large bowel proce	1	-1.5323	0.0937	267.5751	<.0001
mdrg_603	Rectal resection	1	-1.4226	0.1261	127.2062	<.0001
mdrg_604	Peritoneal adhesiolysis	1	-1.9721	0.1107	317.3860	<.0001
mdrg_605	Appendectomy w complicated principal diag	1	-2.4510	0.1421	297.3977	<.0001
mdrg_606	Appendectomy w/o complicated principal diag	1	-2.5886	0.1202	463.5619	<.0001
mdrg_607	Minor small & large bowel procedures	1	-1.8936	0.1441	172.6848	<.0001
mdrg_608	Anal & stomal procedures	1	-2.6775	0.1681	253.5728	<.0001
mdrg_609	Inguinal & femoral hernia procedures	1	-1.6878	0.1312	165.5040	<.0001
mdrg_610	Hernia procedures except inguinal & femoral	1	-1.7756	0.1094	263.2360	<.0001
mdrg_611	Other digestive system O.R. procedures	1	-1.8101	0.1385	170.6834	<.0001
mdrg_701	Pancreas liver & shunt procedures	1	-1.6282	0.1035	247.5026	<.0001
mdrg_702	Biliary tract proc except only cholecyst	1	-1.4123	0.1669	71.6312	<.0001
mdrg_703	Cholecystectomy w c.d.e.	1	-1.9679	0.2720	52.3533	<.0001
mdrg_704	Cholecystectomy except by laparoscope	1	-1.4984	0.1072	195.2174	<.0001
mdrg_705	Laparoscopic cholecystectomy	1	-2.8511	0.0936	928.8070	<.0001
mdrg_706	Hepatobiliary diagnostic procedures	1	-1.9483	0.2647	54.1792	<.0001

Parameter	Label	DF	Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square
mdrg_707	Other hepatobiliary or pancreas procedures	1	-2.1259	0.2572	68.3473	<.0001
mdrg_7701	Heart transplant or implant heart assist sys	1	0.5703	0.0861	43.8546	<.0001
mdrg_801	Combined anterior/posterior spinal fusion	1	0.6735	0.1362	24.4423	<.0001
mdrg_802	Spinal fus exc cerv w spinal curv/malig/infec	1	0.8548	0.1562	29.9532	<.0001
mdrg_803	Spinal fusion except cervical	1	0.1184	0.1090	1.1808	0.2772
mdrg_804	Bilateral or multiple major joint procs	1	-1.8094	0.3311	29.8567	<.0001
mdrg_805	Wnd debrid & skn grft exc hand for musculo	1	0.6728	0.1268	28.1698	<.0001
mdrg_806	Revision of hip or knee replacement	1	-0.6651	0.1444	21.2162	<.0001
mdrg_807	Major joint replacement or reattachment	1	-1.9988	0.1156	299.1545	<.0001
mdrg_808	Cervical spinal fusion	1	0.0207	0.1141	0.0330	0.8558
mdrg_809	Amputation for musculoskeletal sys	1	0.3946	0.1707	5.3439	0.0208
mdrg_810	Biopsies of musculoskeletal system	1	-0.8033	0.2394	11.2578	0.0008
mdrg_811	Hip & femur procedures except major joint	1	-1.1828	0.1321	80.1740	<.0001
mdrg_812	Major joint & limb reattachment	1	-1.4620	0.1976	54.7388	<.0001
mdrg_813	Knee procedures w pdx of infection	1	-1.1045	0.3670	9.0568	0.0026
mdrg_814	Knee procedures w/o pdx of infection	1	-1.5706	0.3475	20.4313	<.0001
mdrg_815	Back & neck proc exc spinal fusion	1	0.4785	0.1148	17.3773	<.0001
mdrg_816	Lower extrem & humer proc	1	-1.8377	0.1756	109.5077	<.0001
mdrg_817	Local excision & removal int fix devices	1	-0.3096	0.1813	2.9162	0.0877
mdrg_819	Soft tissue procedures	1	0.1408	0.1497	0.8842	0.3471
mdrg_820	Foot procedures	1	-1.0981	0.3050	12.9652	0.0003
mdrg_826	Other musculoskelet sys & conn tiss proc	1	-0.3781	0.1677	5.0832	0.0242
mdrg_8899	Non-Extensive O.R. Proc Unrelated to PDX	1	-2.3524	0.0727	1046.9167	<.0001
mdrg_901	Skin graft &/or debrid for	1	-1.8107	0.1213	222.8052	<.0001

Parameter	Label	DF	Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square
	skn ulcer or cellulitis					
mdrg_902	Skin graft &/or debrid exc for skin ulcer	1	-0.3486	0.1096	10.1143	0.0015
mdrg_903	Other skin subcut tiss & breast	1	-0.1895	0.0615	9.4882	0.0021
c-statistic = .769						

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

Not applicable (see above)

2b4.5. 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*)

This analysis evaluates how strongly the risk adjustment model is associated with the event of interest. The measure of discrimination, how well the risk adjustment model distinguishes events from non-events, is the c-statistic (also known as the area under a receiver operating characteristic curve). The c-statistic is computed by assigning each observation a predicted probability of the outcome from the risk-adjustment model, based on the value of the observed covariates and the parameter estimates from the risk-adjustment model. Two copies of the dataset are sorted, first from highest to lowest predicted probability and second from lowest to highest predicted probability. Random sampling is used to create a set of paired observations. Pairs that consist of one event and one non-event (discordant pairs) are kept and concordant pairs are discarded. The c-statistic is a measure of the proportion of discordant pairs of observations for which the observation with the event had a higher predicted probability from the risk-adjustment model than the observation without the event. C-statistics above 0.70 and below 0.80 have moderate discrimination. Above 0.80 the discrimination is high. We did not employ common "goodness of fit" tests because these tests tend to be uninformative with large samples.

We also evaluated the calibration of the risk adjustment model by evaluating how closely observed and predicted rates compare across deciles of the predicted rate. This analysis splits the sample into deciles based on predicted rates, and then compares these rates with the observed rates for the population in each decile. A well calibrated model, or one that does not over or under-estimate risk, will have comparable observed and predicted rates across the risk spectrum.

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

If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

Table 6. Risk adjustment Model Discrimination and Calibration for PSI 09 Perioperative Hemorrhage or Hematoma Rate, per 1,000 in 2-year pooled data (2012-2013)

Predicted Rate Decile	Number of Discharges per Decile	Predicted Rate (per 1,000 surgical discharges)	Observed Rate (per 1,000 surgical discharges)	Observed to Predicted Ratio
1 (lowest)	1,078,040	0.2915	0.3423	1.1741
2	1,078,041	0.4107	0.4731	1.1520
3	1,078,041	1.1033	1.1502	1.0425
4	1,078,041	1.8097	1.7847	0.9862
5	1,078,040	2.3934	2.4155	1.0092
6	1,078,041	3.0988	3.1771	1.0252
7	1,078,041	4.3999	4.5054	1.0240
8	1,078,041	5.9946	6.2576	1.0439
9	1,078,041	8.1434	8.3244	1.0222
10 (highest)	1,078,040	16.5800	16.7100	1.0078
C-Statistic	0.7689			

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2012 - 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov (AHRQ QI Software Version 6.0)

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):
See Table 6 in 2b4.6

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:
See calibration by decile in Table 6 in 2b4.6

2b4.9. Results of Risk Stratification Analysis:
Not applicable

2b4.10. 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 risk-adjustment model has moderately high discrimination, based on a c statistic of 0.7689 (i.e., in 77% of randomly selected pairs of discordant observations, the patient who experienced PSI 09 had a higher probability of experiencing the event than the patient who did not). A model that is well calibrated will have observed values similar to predicted values across the predicted value deciles. This indicator is well calibrated and has good discrimination, as the observed to predicted values across the deciles range between 1.17– 1.01 for all deciles.

2b4.11. 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*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. 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)*

This analysis assesses the probability that a hospital is higher or lower than a benchmark or threshold, given hospital size. It reflects whether the indicator can discriminate the best performing hospitals from the lower performing hospitals.

For this analysis, "benchmark" refers to the smoothed indicator rate based on the 20th percentile of the reference population (i.e., 20% of hospitals have a lower mortality rate or better performance). "Threshold" refers to the indicator rate based on the 80th percentile (i.e., 80% have lower mortality or better performance). Assuming an underlying Gamma distribution for the smoothed rates of the measure, the benchmark and threshold values are identified using population reference rates and signal variances computed from the entire AHRQ QI POA Reference Population. Hospital-level 90% confidence limits for smoothed rates are also computed from the Gamma distribution.

The analysis is reported by size decile, based on the denominator cases, demonstrating performance across hospitals of various sizes. Each hospital is assumed to have an underlying distribution of smoothed rates that follows a Gamma distribution. The parameters of a Gamma distribution are shape and scale. For each hospital the shape is calculated as ((smoothed rate)^2/ smoothed rate variance), and the scale is calculated as (smoothed rate variance / smoothed rate). The smoothed rate variance (aka posterior variance) is calculated as the signal variance – (reliability weight * signal variance). The reliability weight is calculated as (signal variance / (signal variance + noise variance)). Hospitals are ranked by size and grouped into 10 equal categories of size (deciles). The Benchmark and Threshold are compared to the Gamma distribution of the smoothed rates for each hospital to determine if the hospital rate is better or worse than the Benchmark and Threshold rates with 95% probability. This provides a 95% confidence interval for the Benchmark and Threshold rate.

Table 7 reports the proportion of hospitals above (better than) and below (worse than) the Benchmark and Threshold rates and the proportion not classified as either above or below. The proportion of hospitals not classified as either better or worse have rates that fall within the 95% confidence interval.

2b5.2. 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 7. Performance Categories by Hospital Size Decile for PSI 09 Perioperative Hemorrhage or Hematoma Rate Using 2-year pooled data (2012-2013)

Size Decile	No. of Hospitals	Avg. No. of Denominator Discharges Per Hospital	Benchmark			Threshold		
			Proportion Better	Proportion Worse	Proportion Unclassified	Proportion Better	Proportion Worse	Proportion Unclassified
1 (smallest)	357	43.9	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000
2	357	219.7	0.0000	0.0308	0.9692	0.0000	0.0000	1.0000
3	358	521.4	0.0000	0.1089	0.8911	0.0028	0.0028	0.9944
4	357	957.2	0.0000	0.1653	0.8347	0.0336	0.0028	0.9636
5	357	1,533.5	0.0000	0.2437	0.7563	0.1625	0.0056	0.8319
6	358	2,257.4	0.0000	0.3017	0.6983	0.2514	0.0140	0.7346

Size Decile	No. of Hospitals	Avg. No. of Denominator Discharges Per Hospital	Benchmark			Threshold		
			Proportion Better	Proportion Worse	Proportion Unclassified	Proportion Better	Proportion Worse	Proportion Unclassified
7	357	3,131.7	0.0000	0.3557	0.6443	0.3417	0.0084	0.6499
8	358	4,308.4	0.0028	0.4218	0.5754	0.3799	0.0140	0.6061
9	357	6,202.5	0.0028	0.5126	0.4846	0.4818	0.0364	0.4818
10 (largest)	357	11,001.4	0.0056	0.6779	0.3165	0.4846	0.0504	0.4650
Overall	3,573	3,017.2	0.0011	0.2818	0.7170	0.2138	0.0134	0.7727

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2012 - 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov (AHRQ QI Software Version 6.0)

2b5.3. 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?*)

Over all hospitals, this indicator has modest discrimination for identifying low or high performing hospitals; 23% of hospitals can be classified as better or worse than the threshold (the percentage classified as either above or below the threshold) and 28% better or worse than the benchmark (the percentage classified as either above or below the benchmark). However, as hospital size increases, the discrimination increases such that for hospitals in the largest 2 deciles the algorithm classifies 52% - 53% of hospitals against the threshold and 52%-68% of hospitals against the benchmark.

2b6. 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 SDS 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 SDS 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.***

Not applicable

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Not applicable

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

Not applicable

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.e., *what do the results mean*

and what are the norms for the test conducted)
Not applicable

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

The AHRQ QIs use frequently reported administrative data variables. PSI 09 excludes cases with missing discharge disposition, age, sex, discharge quarter, discharge year, and principal diagnosis. These variables are required for indicator construction and are required of all hospital discharge records. The frequency of missing data for each variable is available by state and year from the AHRQ HCUP website (http://www.hcup-us.ahrq.gov/cdstats/cdstats_search.jsp).

2b7.2. 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*)

For these variables, frequencies of missing data are typically less than 1%. It is unlikely that bias would occur from such a low frequency of missing data.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i*.e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

Exclusion of cases with missing data for these key variables is appropriate.

3. Feasibility

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

3a. Byproduct of Care Processes

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

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)
If other:

3b. Electronic Sources

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

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*
ALL data elements are in defined fields in electronic claims

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

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.**
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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
Because the indicator is based on readily available administrative billing and claims data, feasibility is not an issue. This version of the indicator requires present-on-admission (POA) data for risk-adjustment and for specification of the numerator and denominator. POA indicators were added as data elements to the uniform bill form (UB-04) effective October 1, 2007. Hospitals incurred a payment penalty for not including POA status on Medicare records beginning October 1, 2008. Each of the secondary diagnoses in a discharge record can be flagged as "present at the time the order for inpatient admission occurs" or not (see http://www.cdc.gov/nchs/icd/icd9cm_addenda_guidelines.htm). The number of states reporting consistent POA has increased dramatically since 2008.

The AHRQ QI software has been publicly available at no cost since 2001; Users have over ten years of experience using the AHRQ QI software in SAS and Windows.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
There are no fees. Software is freely available from the AHRQ Quality Indicators website (http://www.qualityindicators.ahrq.gov/).

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.

Planned	Current Use (for current use provide URL)
	Public Reporting Arizona Department of Health Services, AZ Hospital Compare, MONAHRQ website http://pub.azdhs.gov/hospital-discharge-stats/2012/AboutQualityRatings.html Cigna http://www.cigna.com/pdf/CentersOfExcellence.pdf Connecticut Department of Health Services, CT Hospital Compare, MONAHRQ website http://ctmonahrq.ct.gov/2012/index.html#/resources/AboutQualityRatings Payment Program CMS Premier Hospital Quality Incentive Demonstration (2003-2009) https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalPremier.html Quality Improvement with Benchmarking (external benchmarking to multiple organizations) University HealthSystem Consortium/Vizient https://www.vizientinc.com/clinical-analytics-and-benchmarking.htm Quality Improvement (Internal to the specific organization) Greenville Health System, Quality and Safety Report http://www.ghs.org/reportcard

4a.1. For each CURRENT use, checked above, provide:

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

Public Reporting
Arizona Department of Health Services, AZ Hospital Compare, MONAHRQ website
Hospital quality ratings from all hospitals in Arizona
http://pub.azdhs.gov/hospital-discharge-stats/2012/AboutQualityRatings.html

CareChex (Division of Quantros)
Provides comprehensive reports of hospitals to consumers, providers and purchasers
http://www.carechex.com/QualityIndicators.aspx

Cigna
Centers of Excellence Hospital Value Tool – Health insurance company
http://www.cigna.com/pdf/CentersOfExcellence.pdf

Connecticut Department of Health Services, CT Hospital Compare, MONAHRQ website
Hospital quality ratings from all hospitals in Connecticut
http://ctmonahrq.ct.gov/2012/index.html#/resources/AboutQualityRatings

Healthcare Association of New York State
Supports availability of hospital quality and safety information to help patients make choices and assist providers in improving care
http://www.hanys.org/quality/data/report_cards/2013/docs/2013_hanys_report_card_book.pdf

HealthGrades
Healthgrades measures 40 million patient records from 4,500 hospitals nationwide for the most recent three-year period. Consumer-targeted hospital and provider ratings

https://d2dcgio3q2u5fb.cloudfront.net/54/98/f79cdfd84640a03792ea092f20a8/2014-patient-safety-methodology.pdf

Illinois State Government
Illinois Hospital Report Card and Consumer Guide to Health Care
http://www.healthcarereportcard.illinois.gov/glossaries/index/#PostoperativeHemorrhageorHematoma

Iowa Healthcare Collaborative
Hospital quality ratings from hospitals in Iowa
https://iowareport.ihconline.org/Public/Reports.aspx?FID=778&F1ID=0&F2ID=0&F3ID=0&CID=2&PID=4

Kentucky Hospital Association Quality Data
Hospital quality ratings from most hospitals in Kentucky
http://info.kyha.com/QualityData/

Maine Health Data Organization (MHDO), MONAHRQ Website
Hospital quality ratings from all hospitals in Maine
https://mhdo.maine.gov/monahrq/#/resources/AboutQualityRatings

Nevada Compare Care, MONAHRQ website
Hospital quality ratings from most hospitals in Nevada
http://nevadacomparecare.net/MQ2014/index.html#/professional/resources/AboutQualityRatings

Nevada Hospital Association
Transparency and Performance of Nevada hospitals for specific clinical indicators
http://www.nvhospitalquality.net/old-home

New Jersey Department of Health
Public report of PSI performance for New Jersey Hospital
http://web.doh.state.nj.us/apps2/hpr/docs/2012/technicalreport_psi.pdf

Niagara Health Quality Coalition, New York State Hospital Report Card
Consumer focused public report of quality indicator performance for NY hospitals.
http://www.myhealthfinder.com/newyork15/main_byproc.php

Norton Healthcare
Report patient satisfaction scores in Norton Healthcare hospitals and their performance on nationally recognized quality indicators and practices http://www.nortonhealthcare.com/QualityReport

Oklahoma State Department of Health, MONAHRQ
Compares quality ratings on hospitals across Oklahoma
https://www.phin.state.ok.us/ahrq/MONAHRQ%202010/Methodology.html

Texas Health Resources
Provides quality and safety reports for all Texas Health Resources
https://www.texashealth.org/Documents/System/Quality_Patient_Safety/Reports/03-02-2016_Surgery.pdf

U.S. News and World Report
National publication that lists ratings of U.S. medical centers based on performance
http://www.usnews.com/pubfiles/BH2015-16MethodologyReport.pdf

Virginia Health Information
Compares quality ratings on hospitals across Virginia
http://www.vhi.org/MONAHRQ/default.asp?yr=2013

Washington State, MONAHRQ website

Information system of inpatient care utilization, quality, and potentially avoidable stays in Washington State's community hospitals
http://www.wamonahrq.net/MONAHRQ_5p0_WA_2012/index.html#/resources/AboutQualityRatings

WHA Information Center (Wisconsin Hospital Association)
Wisconsin Inpatient Hospital Quality Indicators Report
http://www.whainfocenter.com/uploads/PDFs/Publications/QualityIndicators/2012_WI_IQIReport.pdf

Quality Improvement (Internal to the specific organization)
Greenville Health System
Data collected from four hospitals in Greenville Health System, compared with internal rates
http://www.ghs.org/reportcard

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

4a.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.*)
N/A

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

See Table 1 in response to question 1b.2 for 2-year pooled rates (also included in supplemental materials). Additional data discussed below.

Rates of this measure have decreased slightly during 2011-2013 from 4.9 to 4.4 cases/1,000 hospitalizations. This may reflect improvements in care or motivation of providers to adjust documentation and coding practices to minimize the use of the perioperative hemorrhage and hematoma diagnosis codes.

4b.2. **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.**

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
No evidence has been identified suggesting unintended consequences for this measure.

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

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 completely harmonized?

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

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:** PSI_09_Supplemental_files_160526.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Agency for Healthcare Research and Quality
Co.2 **Point of Contact:** Pam, Owens, Pam.Owens@ahrq.hhs.gov, 301-427-1412-
Co.3 **Measure Developer if different from Measure Steward:** Agency for Healthcare Research and Quality
Co.4 **Point of Contact:** Mamatha, Pancholi, Mamatha.Pancholi@ahrq.hhs.gov, 301-427-1470-

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.
In 2002, two workgroups were convened to provide feedback on key indicator development decisions and methodology, including the usefulness of Perioperative Hemorrhage or Hematoma Rate (PSI 09), formerly known as Postoperative Hemorrhage or Hematoma Rate (PSI 09). These workgroups included a multispecialty panel and a surgical specialty panel; the active members were:

Charles Bethea, MD, Cardiologist
Oklahoma City, OK
Duke Clinical Research Institute
Nominated by the American College of Cardiology

John Hunt, MD, MPH, Trauma surgeon, critical care
New Orleans, LA
Health Science Center - Louisiana State University
Nominated by the American College of Surgeons

Franco Laghi, MD, Critical care physician
Maywood, IL
Loyola University
Nominated by the American Thoracic Society

John Nelson, MD, FACP, Internist/Hospitalist
Bellevue, WA
Overlake Hospital Medical Center
Nominated by the National Association of Inpatient Physicians

Carol A. Petersen, RN, BSN, MAOM, CNOR, Perioperative nursing specialist
Denver, CO
Center for Nursing Practice
Nominated by the Association of Peri-Operative Registered Nurses

Bruce Williams, MSN, RN, Critical care nurse specialist
Orangeburg, SC
The Regional Medical Center - of Orangeburg and Calhoun Counties
Nominated by the American Association of Critical-Care Nurses

Preston Winters, MD, FACP, Internist
White Plains, NY
White Plains Hospital Center
Nominated by the American College of Physicians

Rodney Appell, MD, Female urologist
Houston, TX
Baylor College of Medicine

Nominated by the American Urologic Association

Alan Freeland, MD, Orthopedic surgeon
Jackson, MS
University of Mississippi Medical Center
Nominated by the American Academy of Hand Surgeon)

Patricia Howson, MD, MSc, Orthopedic surgeon
Redwood City, CA
Kaiser Permanente
Nominated by the American Academy of Orthopedic Surgeons

William Hozak, MD, Orthopedic surgeon
Philadelphia, PA
Jefferson Medical School
Nominated by the American Association of Hip and Knee Surgeons

Mathew Indeck, MD, General Surgeon -trauma surgery
Danville, PA
Jefferson College of Medicine
Nominated by the American College of Surgeons

Bruce Kaufman, MD, Pediatric neurosurgeon
Milwaukee, WI
Medical College of Wisconsin
Nominated by the American Association of Neurological Surgeons

In 2013, ten panels of experts were convened to support the process of converting the AHRQ QIs from ICD-9-CM to ICD-10-CM/PCS in an accurate and transparent manner, to improve the validity and usefulness of the QIs. One of these panels –focused on general surgical conditions - advised AHRQ on the ICD-10-CM/PCS specifications for PSI 09. The active members of this panel were:

Joel V. Brill, MD, AGAF
Bethesda, MD
AGA Digestive Health Outcomes Registry
Fair Health, Inc.

John Maa, MD
San Francisco, CA
UCSF Dept of Surgery

Richard Dutton, MD, MBA
Park Ridge, IL
Anesthesia Quality Institute

Robert S. Gold, MD
Atlanta, GA
CEO
DCBA, Inc

Lou Ann Schraffenberger, MBA, RHIA, CCS, CCS-P
Oak Brook, Illinois
AHIMA Approved ICD-10-CM/PCS Trainer
Advocate Health Care

Monica VanSuch, MBA, RHIA

Rochester, MN
Division of Health Care Policy and Research
Mayo Clinic

Irene Lopez, BSN, RN,CSTR
Austin, TX
Trauma Services Administration
University Medical Center Brackenridge

Karen Snyder, BSN, RN
Cleveland, Ohio
Cleveland Clinic

Tina Hernandez-Boussard, PhD, MPH
Palo Alto, CA
Stanford University School of Medicine
Division of General Surgery

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2002

Ad.3 **Month and Year of most recent revision:** 06, 2016

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?** 06, 2016

Ad.7 **Disclaimers:** None

Ad.8 **Additional Information/Comments:** None

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 2940
De.2. **Measure Title:** Use of Opioids at High Dosage in Persons Without Cancer
Co.1.1. **Measure Steward:** PQA
De.3. **Brief Description of Measure:** The proportion (XX out of 1,000) of individuals without cancer receiving prescriptions for opioids with a daily dosage greater than 120mg morphine equivalent dose (MED) for 90 consecutive days or longer.
1b.1. **Developer Rationale:** Abuse and overdose of prescription drugs is a major public health issue in the United States.(1,2) Though there is no FDA maximum dose or duration for opioid drugs, studies have demonstrated that patient populations taking high opioid doses for prolonged periods are often characterized by high rates of psychiatric and substance abuse disorders, frequently do not receive care consistent with clinical guidelines, and have higher death rates.(3-6)

PQA developed 3 measures related to prescription opioid use that are indicative of the quality of care for patients taking these medications. The measures examine the quality of use related to the dose of the medications over time, access to the medications through multiple providers, and the combination of both these criteria. This measure, Use of Opioids at High Dosage in Persons Without Cancer, focuses specifically on the use of opioids at high dosage.

Claims data from commercially insured patients indicate that approximately 8% of opioid prescriptions for acute pain and 12% for chronic pain specify a daily dosage of 120mg MED or more.(2) The Washington State Agency Medical Directors Group has suggested 120mg MED as a dosage level that should not be exceeded without special consideration.(4) Group Health Cooperative (GHC), which implemented this guidance from the 2010 edition, has demonstrated a reduction in opioid doses for their patients with chronic pain. For the last quarter of 2014, less than one-quarter of these patients seen by GHC providers received 50 mg/day MED or greater and only 7.3% exceeded 120 mg/day MED.(4) The proportion of patients being treated at this dosage for more than 90 days has not been described. However, one study of veterans treated with 180mg MED/day or more for 90+ days (3) found that this group was characterized by high rates of psychiatric and substance abuse disorders and frequently did not receive care consistent with clinical guidelines. Studies suggest that high opioid dosage increases the risk of overdoses and fractures.(5-7)

Data suggest that efforts to prevent opioid overdose deaths should include a multi-faceted approach focused on strategies that target high-dose opioid users as well as persons who seek care from multiple doctors and pharmacies. The data also suggest that these criteria can be considered separately, as measures related to prescribed opioids for appropriate clinical uses versus inappropriate uses. Thus, as stated above, PQA developed 3 measures: one for high dose therapy, one for multiple providers, and one that is the intersection of both high dose and multiple providers – with this measure presently under consideration focused specifically on the use of opioids at high dosage.

References:
1. U.S. Department of Health and Human Services, Office of Disease Prevention and Health Promotion. (2014). National Action Plan for Adverse Drug Event Prevention. Washington, DC. Accessed on: 4/9/15. Available at: http://www.health.gov/hcq/pdfs/ADE-Action-Plan-508c.pdf.
2. Liu Y, Logan JE, Paulozzi LJ, et al. Potential misuse and inappropriate prescription practices involving opioid analgesics. Am J Manag Care. 2013;19:648-65. PMID: 24304213.
3. Morasco BJ, Duckart JP, Carr TP, et al. Clinical characteristics of veterans prescribed high doses of opioid medications for chronic non-cancer pain. Pain. 2010;151:625-32. PMID: 20801580.
4. Agency Medical Directors Group (AMDG). Interagency Guideline on Opioid Dosing for Chronic Non-cancer Pain: An educational aid to improve care and safety with opioid therapy. 2010 Update. Accessed on: 4/9/15. Available at: http://www.agencymeddirectors.wa.gov/files/opioidgdline.pdf.
5. Dunn KM, Saunders KW, Rutter CM, et al. Opioid prescriptions for chronic pain and overdose: a cohort study. Ann Intern Med.

2010;152:85-92. PMID: 20083827.
6. Paulozzi LJ, Kilbourne EM, Shah NG, et al. A history of being prescribed controlled substances and risk of drug overdose death. Pain Med. 2012;13:87-95. PMID: 22026451.
7. Saunders KW, Dunn KM, Merrill JO, et al. Relationship of opioid use and dosage levels to fractures in older chronic pain patients. J Gen Intern Med. 2010;25:310-5. PMID: 20049546.

S.4. **Numerator Statement:** Any member in the denominator with opioid prescription claims where the MED is greater than 120mg for 90 consecutive days or longer*

*MED calculation is included in S.6 Numerator Details

S.7. **Denominator Statement:** Any member with two or more prescription claims for opioids filled on at least two separate days, for which the sum of the days supply is greater than or equal to 15.

S.10. **Denominator Exclusions:** Any member with a diagnosis for Cancer or a Prescription Drug Hierarchical Condition Category (RxHCC) 8, 9, 10, or 11 for Payment Year 2015; or RxHCC 15, 16, 17, 18, or 19 for Payment Year 2016 (see list in S.11 and S.2b); or a hospice indicator (Medicare Part D) from the enrollment database.

De.1. **Measure Type:** Process

S.23. **Data Source:** Administrative claims

S.26. **Level of Analysis:** Health Plan, Population : National, Population : State

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

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

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

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

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *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.

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 benefits of high-dose opioids for chronic pain are not established and the risks for serious harms related to opioid therapy increase at higher opioid dosage. Higher opioid dosages are associated with increased risks for motor vehicle injury, opioid use disorder, and overdose. The risk for overdose increases in a dose-dependent manner. Lower dosages of opioids reduce the risk for overdose, but a single dosage threshold for safe opioid use has not been identified.

Exception to evidence

N/A

Guidance from the Evidence Algorithm

1-No → 3-Yes→ 4-Yes → 5a-Yes→HIGH

Questions for the Committee:

- *What is the relationship of this measure to patient outcomes?*
- *How strong is the evidence for this relationship?*
- *Is the evidence directly applicable to the process of care being measured?*

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

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

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

- The measure was tested in three different health plan data sources – the Medicare population, one commercial heath plan, and the Medicaid population.
- The testing from the Medicare population used administrative claims data from January 1st 2013 to December 31st 2013. The Medicare rates ranged from 30.0 per 1,000 to 49.66 per 1,000. The Mean was 39.27 per 1,000 and the median was 38.7 per 1,000. The standard deviation was 8.32. The 25th percentile was 34.62 per 1,000, the 50th percentile is the median (38.70 per 1,000) and the 75th percentile was 43.35 per 1,000. The interquartile range was 8.73.
- The Medicaid rates ranged from 8.15 per 1,000 to 66.45 per 1,000. The Mean was 34.04 per 1,000 and the median was 34.29 per 1,000. The standard deviation was 20.61. The 25th percentile was 20.4 per 1,000, the 50th percentile is the median (34.29 per 1,000) and the 75th percentile was 48.1 per 1,000. The interquartile range was 27.68.
- Testing was also conducted in one Commercial health plan using administrative claims from January 1st 2013 to December 31st 2013. This plan covered 209,191 individuals age 18 and older. The measure rate for this plan was 32.03 per 1,000.

Disparities

- The beneficiary level Low Income Subsidy (LIS) variable was used to determine disparities in rates for populations with different sociodemographic status. The LIS is a subsidy paid by the Federal government to the drug plan for Medicare beneficiaries who need extra help with their prescription drug costs due to limited income and resources. The measure rate for the LIS group is 62.41 per 1,000 while the rate for the non-LIS population is significantly lower, at 28.09 per 1,000.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*

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:

Comments:**This is a process measure using administrative claims data. A systematic review of the evidence is provided.
**This is a timely measure given opiate abuse and negative outcomes

1b. Performance Gap:
Comments:
**Testing of the measure was done in 3 health plan resources - Medicare, one commercial plan and Medicaid for Jan. 2013 to Dec. 2013. The rates per 1000 for each plan indicated an opportunity to improve performance. There is a significant disparity noted between beneficiaries receiving the LIS and those who do not.

**Clearly gap and developer provides evidence of disparity

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Administrative claims, health plan enrollment information

Specifications:

- This measure assesses the proportion of individuals without cancer receiving prescriptions for opioids with a daily dosage greater than 120mg morphine equivalent dose (MED) for 90 consecutive days or longer.
- The level of analysis (i.e., the measured entity) is the prescription drug health plan.
    - The developer notes that the measure also contains claims data from multiple care settings, including ambulatory, skilled nursing facility, pharmacy etc.
    - The measure is stratified by the following lines of business for the health plan:
        - Commercial
        - Medicare
        - Medicaid
- The measure is reported as a rate (per 1,000 plan members).
- The measure uses health plan medical and pharmacy claims and health plan member enrollment information as its data sources.
- To identify the denominator population, the measure identifies any member with two or more prescription claims for opioids filled on at least two separate days, for which the sum of the days' supply is greater than or equal to 15.
- To derive the numerator, the measure calculates the daily MED of opioid claims for each member and identifies the days where the MED threshold (120 MEDs) is exceeded; any member for whom the MED threshold is exceeded for at least 90 consecutive days is included in the numerator.
- A list of opioid medications (along with the MED conversion factor) is provided in the submission form.
- The measure excludes patients with a diagnosis of cancer and patients in hospice.
- A list of administrative codes (ICD-9/10, RxHCC) identifying denominator exclusions is provided in a spreadsheet attached to the measure submission.

Questions for the Committee :

- *Specific questions on the specifications, codes, definitions, etc.*
- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

SUMMARY OF TESTING

Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**
Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

- The developer used several data sets for reliability testing:
  - For Medicare testing, the analysis included a convenience sample of over 700 Medicare Part D prescription drug plans (comprising a total of 7,067,445 individuals aged 18 and older)
  - Testing was also conducted in one Commercial health plan (comprising a total of 209,191 individuals age 18 and older)
  - For Medicaid testing, the analysis included 8 state-based prescription drug plans covering 6 states (comprising a total of 1,437,410 individuals age 18 and older)

Method(s) of reliability testing

- To demonstrate reliability, the developer conducted a signal-to-noise analysis of the computed measure score using a beta-binomial model.
  - The developer explains that a reliability score (i.e., signal-to-noise ratio) may range from 0 to 1; a score of 0 signifies that all variation is due to measurement error ("noise"), while a score of 1 signifies that all variation represents true differences in performance scores between plans ("signal").

Results of reliability testing

- The developer provides the results of reliability testing in a table presenting the distribution of individual plan reliability scores; the mean reliability score across all plans is **0.9938**.
- The developer suggests that a reliability score of 0.7 is the minimum threshold for reliability, and that based on the high scores achieved in the analysis, this measure should be considered reliable.

Questions for the Committee:
- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

Guidance from the Reliability Algorithm

[Box 1] Specifications precise and unambiguous → [Box 2] Empirical testing conducted on the measure as specified → [Box 4] Testing conducted at the measure score level → [Box 5] → Testing method described and appropriate → [Box 6] High certainty or confidence that measure scores are reliable → [Box 6a]

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

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:
- *Are the specifications consistent with the evidence?*

2b2. Validity testing

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.

SUMMARY OF TESTING

Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

- ☒ **Face validity only**
- ☐ **Empirical validity testing of the measure score**

Validity testing method:

- To demonstrate validity, the developer cites their (PQA's) approach to measure development and testing.
    - This approach includes identification of important concepts by PQA member workgroups, evaluation and refinement of concepts by the PQA Quality Metrics Expert Panel (QMEP), partnership with measure development experts, and processes for review, comment, and approval by PQA members.
- The developer notes that the QMEP Panel reviewed the results of measure testing, including performance measure scores, and provided an assessment of whether measure results reflect quality of care.

Validity testing results:

- The developer reports that out of 12 QMEP members voting on the measure's face validity, 67 percent strongly agreed that the measure results reflected quality of care.
- In addition, the developer notes that of 89 PQA members voting on whether to endorse the measure, 69.7 voted in favor of approval.
- Five PQA member organizations also tested the measure using their own data, and all strongly agreed that the measure reflected the quality of care provided for their populations.

- ***NQF Staff Note:*** *Assessment of this measure's validity appears to have been conducted by the same groups involved in development of the measure; NQF prefers face validity to be assessed by experts or other stakeholder groups who have not been involved in the measure development process.*

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- This measure excludes patients with a diagnosis of cancer and patients in hospice.
- The developer's rationale for these exclusions is that patients at end of life, undergoing hospice care, and those with cancer may have unusual requirements for pain management; the developer notes that these exclusions are consistent with the 2016 CDC Guideline for Prescribing Opioids for Chronic Pain.
- Because prescription claims data do not contain claims for palliative medication, such as opioids, for persons in Medicare Part D that are in hospice care, this exclusion was not tested by the developer. In addition, for the Medicaid population, the majority of the plans were not able to identify hospice exclusions in their data.
- For the cancer exclusion, the developer provided an analysis of data from eight health plans, identifying the number of exclusions and the percent of the overall population that would be affected by including patients with cancer diagnoses.
- The developer reports that the cancer patient exclusions were 0.5% to 1.9% of the overall population. The one Medicaid plan that could identify hospice exclusions found only 15 cases, which represented 0.003% of their total population.
- Interpreting the results of this analysis, the developer states that in some plans, almost 2% of the population has cancer and would be included in the measure if cancer was not excluded, suggesting that this is a significant proportion of the population that could potentially impact the measure rates.

- The developer states that no inferences about the hospice exclusion could be drawn because the majority of the plans could not identify exclusions.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

Questions for the Committee:

- *Is an appropriate risk-adjustment strategy included in the measure?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- To assess the measure's ability to identify meaningful differences in performance, the developer analyzed their testing data to identify the mean, median, standard deviation, and interquartile range for the measure rates for the Medicare population and the Medicaid population.
- In addition, for the Medicaid population, the rates were divided into quartiles, and a Student's t-test was used to compare the rates between the bottom quartile (75th percentile) and top quartile (25th percentile).
- For the Medicare population, the developer reports a mean performance rate (per 1,000 members) of **39.27**, a median rate of **38.70**, and a standard deviation of **8.32**.
- For the Medicaid population, the developer reports a mean performance rate (per 1,000 members) of **34.04**, a median rate of **34.29**, and a standard deviation of **20.61**.
- The following tables provides additional results of the developer's analysis:

Medicare Population

Minimum	30.00
25th Percentile	34.62
50th Percentile	38.70
75th Percentile	43.35
Maximum	49.66
Interquartile Range	8.73

Medicaid Population

Minimum	8.15
25th Percentile	20.4
50th Percentile	34.3
75th Percentile	48.1
Maximum	66.45
Interquartile Range	27.68
Student's t-test p-value	0.029

- The developer's interpretation of these results is that the measure rates showed significant variation in the Medicare population, and even greater variation in the Medicaid population
- The developer also states that there is a statistically significant difference in measure rates between the top and bottom quartile of the plans included in the testing (P=0.029 at alpha=0.05), and suggests that this variation shows that there are meaningful differences in rates across plans.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

N/A

2b7. Missing Data

- The developer notes that since all data elements are available via prescription claims data, it is not expected—nor was it found—that missing data would result.
- The developer states that, as a result, performance results would not be biased, as prescription claims data provides the data elements necessary to calculate the measure rate.

Guidance from the Validity Algorithm

[Box 1] Specifications consistent with evidence → [Box 2] Potential threats to validity addressed → [Box 3] Empirical validity testing **NOT** conducted using the measure as specified → [Box 4] Face validity systematically assessed → [Box 5] Results indicate substantial agreement that performance score can be used to distinguish quality → [Moderate]

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

Committee pre-evaluation comments

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

2a1. & 2b1. Specification:

Comments: **Specifications are clear and consistent with evidence.
**Elements clear-uses multiple sources to determine prescriptions

2a2. Reliability Testing

Comments:**Measure score reliability testing was done using a signal to noise ratio. The mean reliability score across all plans ws 0.9938 suggesting high reliability
**Reliability seems surprisingly high

2b2. Validity Testing:

Comments:

- Face validity testing was conducted using the same group that developed the measure. 67% of the panel agreed the measure reflects quality of care, 69.7% voted in favor of approval.
- Meets minimal validity but not tested by any outside groups other than developer

2b3. Exclusions Analysis
2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures
2b5. Identification of Statistically Significant & Meaningful Differences In Performance
2b6. Comparability of Performance Scores When More Than One Set of Specifications
2b7. Missing Data Analysis and Minimizing Bias

Comments:

**Exclusions are supported.
**No risk-adjustment method.
**Measure rates show significant variation in the Medicare population, and even greater variation in the Medicaid population.
**There is a statistically significant difference in measure rates between the top and bottom quartile of the plans included.
exclusions clear no risk adjustment
**Validity moderate

Criterion 3. Feasibility

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.

- This measure is generated or collected by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims) Other data elements include Prescription claims data.
- ALL data elements are in defined fields in electronic claims.
- Pilot test sites indicated the measure was feasible and results were able to be reported efficiently and accurately. CMS calculates the measure for Part D plans. The prescription claims and medical data is readily available.
- Certain uses of the Measures are only approved with a licensing agreement from the developer, that specifies the terms of use and the licensing fee. The developer reserves the right to determine the conditions under which it will approve and/or license the Measures.

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

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments:

**Data is available in electronic claims. Pilot sites indicated the measure was feasible and results were able to be reported efficiently and accurately.

**Seems easy to capture electronically

Criterion 4: Usability and Use

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

Current uses of the measure:

Publicly reported? ☐ Yes ☒ No

Current use in an accountability program? ☒ Yes ☐ No

OR

Planned use in an accountability program? ☐ Yes ☒ No

Accountability program details:

- The measure was developed in 2015.

- The measure is currently being used in the Medicare Part D Overutilization Monitoring System to monitor the utilization of opioids for members with the Medicare drug benefit. CMS Medicare Part D Drug Benefit - Purpose: Monitor Opioid use by Medicare Part D beneficiaries - Geographic area: National, approximately 38 million beneficiaries in Medicare Part D plans.
- CMS has announced plans to move this measure into the 2019 Part D Display Measures, using data from 2017.
- This measure also has been included in the 2016 Medicaid Adult Core Set.
- Reporting of results is not yet available.

Improvement results:

- There are no improvement results, as this is the initial endorsement submission.

Unexpected findings (positive or negative) during implementation:

- Developer did not identify any specific unexpected findings related to this measure.

Potential harms:

- Although no unintended negative consequences to individuals or populations were identified during testing, , concerns have been raised that prescribing changes such as dose reduction (without offering or arranging evidence-based treatment for patients with opioid use disorder) might be associated with unintended negative consequences, such as patients seeking heroin or other illicitly obtained opioids (1,2) or interference with appropriate pain treatment.(3) Data indicate that if access to prescription opioids is limited, some users of opioid analgesics will transition to heroin or other illicitly obtained opioids, leading to increased overdose death coincident with prescribing restrictions.(1) There are also concerns about pain being underdiagnosed and undertreated, particularly for ethnic and racial minorities, females, children or infirm elderly, or individuals who may be perceived to have mental health problems or are drug seeking.(4,5)
- These concerns must be balanced by the current situation in the United States which has been described by the CDC as an epidemic of opioid abuse, overdose, and deaths. Overdose involving opioid analgesics killed almost 17,000 persons in 2010 and the number of people with opioid analgesic use disorders increased to nearly 2 million.(6) The developer believes the potential benefits of monitoring those patients receiving the very highest doses of opioids for extended periods of time or receiving these drugs from multiple providers outweighs potential negative consequences.

Feedback :

- Developer did not identify any specific feedback loops related to this 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

4. Usability and Use

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments:

**Not currently used in public reporting. Currently used in accountability programs for Medicare Part D and the Medicaid Adult Core set.

**Risk/benefit analysis of limiting access to opioid that could lead to illicit drug seeking behavior vs. current US epidemic of opioid abuse.
**Could be used for accountability but current application says not planned to be used for accountability
Since this is a new measure needs to be tested in real world before public reporting

Criterion 5: Related and Competing Measures

Related or competing measures

- The measure is related to 2950 and 2951 which are being proposed for endorsement.

Harmonization

- N/A

Pre-meeting public and member comments

Submitted By: ADVault, Inc.

ADVault believes that people live better lives and, if in a health crisis, can receive better care when they have confidence they can be involved in the creation and implementation of their medical treatment plans and decisions, factors extremely important when it comes to addictive, narcotic medications like opioids. To do so, they must be able to communicate and express their goals, preferences and priorities for care in a meaningful and actionable way so providers can consider those thoughts. At some point in life, everyone will lose his or her ability to communicate effectively and understand what is being asked of him or her. Healthcare agents should have the confidence to know those value statements as well, in order to fulfill their role as surrogate decision-makers. Non-surrogate family members are comforted with third-party decision-making if they have proof the patient's voice is being heard, clearly understood, and to the extent possible, honored.

Therefore, ADVault strongly recommends providers (1) search for a person's digital emergency, critical and advance care plan (ECACP) upon admission and each time the patient is transitioned to a new site of care, (2) review and update the ECACP in various stages of a person's admission (outpatient or inpatient) and/or illness to ensure respect for the person's goals, preferences and priorities for care, (3) link the digital ECACP to the EHR and/or patient portal in order to ease access and address security, privacy and patient consent concerns, (4) track and make available the number of ECACPs found, opened and re-visited, and the impact they have on the care of the patient, as well as patient, family and caregiver satisfaction, such data to be reported in a manner such that: (a) consumers can make better choices about hospitals and doctors; (b) doctors improve the satisfaction and quality of their work; and (c) hospital administrators gauge performance and align caregiving goals with actual outcomes. Finally, if no ECACP can be found via standards-based healthcare IT transport mechanisms, the hospital/provider should engage the patient to create one whenever possible.

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Use of Opioids at High Dosage in Persons Without Cancer
IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: 5/13/2016

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied*

together.
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

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

Outcome

☐ Health outcome: Click here to name the health outcome

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

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

☒ Process: Prescriptions for high doses of opioids

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

It has been shown that the measured process, prescriptions for high-doses of opioids, correlates with undesired health outcomes. Benefits of high-dose opioids for chronic pain are not established and the risks for serious harms related to opioid therapy increase at higher opioid dosage. Higher opioid dosages are associated with increased risks for motor vehicle injury, opioid use disorder, and overdose. The risk for overdose increases in a dose-dependent manner. Lower dosages of opioids reduce the risk for overdose, but a single dosage threshold for safe opioid use has not been identified.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☒ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☒ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

CDC Guideline: Dowell D, Haegerich TM, Chou R. CDC Guideline for Prescribing Opioids for Chronic Pain - United States, 2016. MMWR Recomm Rep. 2016 Mar 18;65(1):1-49. doi: 10.15585/mmwr.rr6501e1. Available at: http://www.cdc.gov/drugoverdose/prescribing/guideline.html.

AMDG Guideline: Interagency Guideline on Prescribing Opioids for Pain: Developed by the Washington State Agency Medical Directors' Group (AMDG) in collaboration with an Expert Advisory Panel, Actively Practicing Providers, Public Stakeholders, and Senior State Officials. June 2015. Available at: www.agencymeddirectors.wa.gov.

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

CDC Guideline: Recommendation 5, pages 22-24. "When opioids are started, clinicians should prescribe the lowest effective dosage. Clinicians should use caution when prescribing opioids at any dosage, should carefully reassess evidence of individual benefits and risks when considering increasing dosage to ≥50 morphine milligram equivalents (MME)/day, and should avoid increasing dosage to ≥90 MME/day or carefully justify a decision to titrate dosage to ≥90 MME/day (recommendation category: A, evidence type: 3)."

AMDG Guideline: Recommendation 3, page 11. "Do not escalate COAT [chronic opioid analgesic therapy] to more than 120 mg/day MED without first obtaining a consultation from a trained pain specialist who agrees that a high dose is indicated and appropriate. Providers must routinely monitor and document sustained improvement in function and

quality of life and an absence of the risk factors listed in recommendations 1 and 2."

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

CDC Guideline: Category A recommendation: Applies to all persons; most patients should receive the recommended course of action.

AMDG Guideline: N/A

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

CDC Guideline: Recommendation Categories
Based on evidence type, balance between desirable and undesirable effects, values and preferences, and resource allocation (cost).
Category A recommendation: Applies to all persons; most patients should receive the recommended course of action.
Category B recommendation: Individual decision making needed; different choices will be appropriate for different patients. Clinicians help patients arrive at a decision consistent with patient values and preferences and specific clinical situations.

AMDG Guideline: N/A

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☒ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

CDC Guideline: Dowell D, Haegerich TM, Chou R. CDC Guideline for Prescribing Opioids for Chronic Pain - United States, 2016. MMWR Recomm Rep. 2016 Mar 18;65(1):1-49. doi: 10.15585/mmwr.rr6501e1. Available at: http://www.cdc.gov/drugoverdose/prescribing/guideline.html.

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

CDC Guideline: The CDC guideline was developed using the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) method (http://www.gradeworkinggroup.org). A previously published systematic review sponsored by the Agency for Healthcare Research and Quality (AHRQ) on the effectiveness and risks of long-term opioid treatment of chronic pain (http://www.effectivehealthcare.ahrq.gov/ehc/products/557/1971/chronic-pain- opioid-treatment-report-141007.pdf, http://dx.doi.org/10.7326/M14-2559) initially served to directly inform the recommendation statements. CDC conducted additional literature searches to update the AHRQ evidence review; more details about the literature search strategies and GRADE methods applied are provided in the Clinical Evidence Review (http://stacks.cdc.gov/view/cdc/38026). CDC developed GRADE evidence tables to illustrate the quality of the evidence for each clinical question. As identified in the AHRQ-sponsored clinical evidence review, the overall evidence base for the effectiveness and risks of long-term opioid therapy is low in quality per the GRADE criteria. Thus, CDC conducted a Contextual Evidence Review (http://stacks.cdc.gov/view/cdc/38027) to provide additional information, including the epidemiology of opioid pain medication overdose. CDC constructed narrative summaries of this contextual evidence and used the information to support the clinical recommendations.

Complete section 1a.7

_____

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

CDD Guideline: The CDC Clinical Evidence Review evaluated the clinical questions regarding the effectiveness, benefits, and harms of long-term opioid therapy (use of opioids on most days for >3 months) for chronic pain. The CDC Contextual Evidence Review focused on the effectiveness of alternative treatments, benefits and harms of opioid therapy; provider and patient values and preferences; and resource allocation.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

CDC Guideline: Type 3 evidence: Observational studies or randomized clinical trials with notable limitations.

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

CDC Guideline: Evidence Type: Based on study design as well as a function of limitations in study design or implementation, imprecision of estimates, variability in findings, indirectness of evidence, publication bias, magnitude of treatment effects, dose-response gradient, and constellation of plausible biases that could change effects.
Type 1 evidence: Randomized clinical trials or overwhelming evidence from observational studies.
Type 2 evidence: Randomized clinical trials with important limitations, or exceptionally strong evidence from observational studies.
Type 3 evidence: Observational studies or randomized clinical trials with notable limitations.
Type 4 evidence: Clinical experience and observations, observational studies with important limitations, or randomized clinical trials with several major limitations.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: January 2008 through August 2014

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

CDC Guideline: In the CDC Guideline, for Key Question 2, specifically related to how harms vary depending on the opioid dose used, 6 observational studies were included from the Clinical Evidence Review (CDC p. 44, Table 1). Five additional observational studies on the association of opioid dosage and overdose risk were identified in the Contextual Evidence Review and considered in the CDC Guideline. These had been excluded from the clinical evidence review because patient samples were not restricted to patients with chronic pain only.

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

CDC Guideline: The overall quality of evidence across studies in the complete body of evidence is low in quality per the GRADE criteria. The relevant studies related to Key Question 2 were described as fair- to good-quality observational studies. These assessments were primarily related to serious study limitations. For risk of overdose related to MME/day, there was no inconsistency or imprecision, and the magnitude of effect and dose response relationship were notable.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

CDC Guideline: Evidence is insufficient (0 studies). The benefits of high-dose opioids for chronic pain are not established.

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

CDC Guideline: The Clinical Evidence Review found that risks for serious harms related to opioid therapy increase at higher opioid dosage. Higher opioid dosages are associated with increased risks for motor vehicle injury, opioid use disorder, and overdose. The Clinical and Contextual Evidence Reviews found that opioid overdose risk increases in a dose-response manner, that dosages of 50–<100 MME/day have been found to increase risks for opioid overdose by factors of 1.9 to 4.6 compared with dosages of 1–<20 MME/day, and that dosages ≥100 MME/day are associated with increased risks of overdose 2.0–8.9 times the risk at 1–<20 MME/day.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

1a.8 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.8.1 What process was used to identify the evidence?

The Centers for Medicare and Medicaid (CMS) Part D Overutilization Monitoring System (OMS) was identified on the CMS website: https://www.cms.gov/Medicare/Prescription-Drug-coverage/PrescriptionDrugCovContra/RxUtilization.html.

1a.8.2. Provide the citation and summary for each piece of evidence.

1. CMS. Medicare Part D Overutilization Monitoring System (OMS) Summaryhttps://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovContra/Downloads/Fact-Sheet-Overutilization-Monitoring-System-11032015.pdf
Summary: CMS developed a comprehensive morphine equivalent dose (MED) approach to assist Part D sponsors in identifying high risk beneficiaries. Beneficiaries who are dispensed opioids that exceed 120 mg of cumulative MED for at least 90 consecutive days, and whose opioid prescriptions are associated with more than 3 prescribers and more than 3 pharmacies are identified as high-risk beneficiaries (i.e., potential opioid overutilizers). This approach was based on the method used in Washington State, as well as the opioid product list and MED conversion factors maintained by the CDC. This cumulative MED approach to identify high risk use of opioids is now being widely adopted outside of Part D.

2. CMS. Advance Notice of Methodological Changes for Calendar Year (CY) 2017 for Medicare Advantage (MA) Capitation Rates, Part C and Part D Payment Policies and 2017 Call Letter. Available at: https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Downloads/Advance2017.pdf
Summary: Part D sponsors have had a significant impact on reducing overutilization of opioids and APAP. From 2011 through 2015, there was a 47% decrease or 13,753 fewer Medicare Part D beneficiaries identified as potential opioid overutilizers (i.e., beneficiaries with at least 90 consecutive days with greater than 120 mg MED daily with more than 3 [i.e., 4 or more] prescribers and more than 3 [i.e., 4 or more] pharmacies contributing to their opioid claims). This represents a 57% decrease in the share of beneficiaries using opioids who are identified as potential opioid overutilizers.

3. Franklin GM, Mai J, Turner J, Sullivan M, Wickizer T, Fulton-Kehoe D. Bending the prescription opioid dosing and mortality curves: impact of the Washington State opioid dosing guideline. *Am J Ind Med.* 2012;55(4):325-31. doi:10.1002/ajim.21998. PMID: 22213274.
In a retrospective observational study using data from WA state workers' compensation system, the 2007 introduction of an opioid dosing guideline in WA appeared to be associated temporally with a 26% decline in the average dose for long-acting opioids and a 35% decline in percent of claimants receiving opioid doses of at least 120 mg MED per day. There was a 50% decrease in opioid-related deaths among injured workers from 2009 to 2010.

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
1_PQA-Opioids_High_Dose_Evidence_Form_051016.docx

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., the benefits or improvements in quality envisioned by use of this measure)*
Abuse and overdose of prescription drugs is a major public health issue in the United States.(1,2) Though there is no FDA maximum dose or duration for opioid drugs, studies have demonstrated that patient populations taking high opioid doses for prolonged periods are often characterized by high rates of psychiatric and substance abuse disorders, frequently do not receive care consistent with clinical guidelines, and have higher death rates.(3-6)

PQA developed 3 measures related to prescription opioid use that are indicative of the quality of care for patients taking these medications. The measures examine the quality of use related to the dose of the medications over time, access to the medications through multiple providers, and the combination of both these criteria. This measure, Use of Opioids at High Dosage in Persons

Without Cancer, focuses specifically on the use of opioids at high dosage.

Claims data from commercially insured patients indicate that approximately 8% of opioid prescriptions for acute pain and 12% for chronic pain specify a daily dosage of 120mg MED or more.(2) The Washington State Agency Medical Directors Group has suggested 120mg MED as a dosage level that should not be exceeded without special consideration.(4) Group Health Cooperative (GHC), which implemented this guidance from the 2010 edition, has demonstrated a reduction in opioid doses for their patients with chronic pain. For the last quarter of 2014, less than one-quarter of these patients seen by GHC providers received 50 mg/day MED or greater and only 7.3% exceeded 120 mg/day MED.(4) The proportion of patients being treated at this dosage for more than 90 days has not been described. However, one study of veterans treated with 180mg MED/day or more for 90+ days (3) found that this group was characterized by high rates of psychiatric and substance abuse disorders and frequently did not receive care consistent with clinical guidelines. Studies suggest that high opioid dosage increases the risk of overdoses and fractures.(5-7)

Data suggest that efforts to prevent opioid overdose deaths should include a multi-faceted approach focused on strategies that target high-dose opioid users as well as persons who seek care from multiple doctors and pharmacies. The data also suggest that these criteria can be considered separately, as measures related to prescribed opioids for appropriate clinical uses versus inappropriate uses. Thus, as stated above, PQA developed 3 measures: one for high dose therapy, one for multiple providers, and one that is the intersection of both high dose and multiple providers – with this measure presently under consideration focused specifically on the use of opioids at high dosage.

References:
1. U.S. Department of Health and Human Services, Office of Disease Prevention and Health Promotion. (2014). National Action Plan for Adverse Drug Event Prevention. Washington, DC. Accessed on: 4/9/15. Available at: http://www.health.gov/hcq/pdfs/ADE-Action-Plan-508c.pdf.
2. Liu Y, Logan JE, Paulozzi LJ, et al. Potential misuse and inappropriate prescription practices involving opioid analgesics. Am J Manag Care. 2013;19:648-65. PMID: 24304213.
3. Morasco BJ, Duckart JP, Carr TP, et al. Clinical characteristics of veterans prescribed high doses of opioid medications for chronic non-cancer pain. Pain. 2010;151:625-32. PMID: 20801580.
4. Agency Medical Directors Group (AMDG). Interagency Guideline on Opioid Dosing for Chronic Non-cancer Pain: An educational aid to improve care and safety with opioid therapy. 2010 Update. Accessed on: 4/9/15. Available at: http://www.agencymeddirectors.wa.gov/files/opioidgdline.pdf.
5. Dunn KM, Saunders KW, Rutter CM, et al. Opioid prescriptions for chronic pain and overdose: a cohort study. Ann Intern Med. 2010;152:85-92. PMID: 20083827.
6. Paulozzi LJ, Kilbourne EM, Shah NG, et al. A history of being prescribed controlled substances and risk of drug overdose death. Pain Med. 2012;13:87-95. PMID: 22026451.
7. Saunders KW, Dunn KM, Merrill JO, et al. Relationship of opioid use and dosage levels to fractures in older chronic pain patients. J Gen Intern Med. 2010;25:310-5. PMID: 20049546.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
The measure was tested in three different health plan data sources – the Medicare population, one commercial heath plan, and the Medicaid population.

The testing from the Medicare population used administrative claims data from January 1st 2013 to December 31st 2013. For the Medicare testing, the analysis included a convenience sample of over 700 Medicare Part D prescription drug plans, covering 7,067,445 individuals aged 18 and older.

The Medicare rates ranged from 30.0 per 1,000 to 49.66 per 1,000. The Mean was 39.27 per 1,000 and the median was 38.7 per 1,000. The standard deviation was 8.32. The 25th percentile was 34.62 per 1,000, the 50th percentile is the median (38.70 per 1,000) and the 75th percentile was 43.35 per 1,000. The interquartile range was 8.73.

The majority of testing used Medicaid prescription claims data from January 1st 2015-December 31st 2015. Testing also included prescription claims data from one state's Medicaid plan from July 1st 2014-June 30th 2015. Testing included 8 state based prescription drug plans in 6 states, covering 1,437,410 individuals age 18 and older.

The Medicaid rates ranged from 8.15 per 1,000 to 66.45 per 1,000. The Mean was 34.04 per 1,000 and the median was 34.29 per

1,000. The standard deviation was 20.61. The 25th percentile was 20.4 per 1,000, the 50th percentile is the median (34.29 per 1,000) and the 75th percentile was 48.1 per 1,000. The interquartile range was 27.68.

Testing was also conducted in one Commercial health plan using administrative claims from January 1st 2013 to December 31st 2013. This plan covered 209,191 individuals age 18 and older. The measure rate for this plan was 32.03 per 1,000.

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*
Disparities data is available for the Medicare population. The testing from the Medicare population used administrative claims data from January 1st 2013 to December 31st 2013 and included a convenience sample of over 700 Medicare Part D prescription drug plans, covering 7,067,445 individuals aged 18 and older.

The beneficiary level Low Income Subsidy (LIS) variable was used to determine disparities in rates for populations with different sociodemographic status. The LIS is a subsidy paid by the Federal government to the drug plan for Medicare beneficiaries who need extra help with their prescription drug costs due to limited income and resources. The measure rate for the LIS group is 62.41 per 1,000 while the rate for the non-LIS population is significantly lower, at 28.09 per 1,000.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.
N/A

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. Demonstrated high priority aspect of healthcare
Affects large numbers, A leading cause of morbidity/mortality, Patient/societal consequences of poor quality
1c.2. If Other:

1c.3. Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.
The misuse of prescription opioids in America is a public health crisis and addressing the overdose epidemic is a high priority for the US government.(1-4) Deaths from drug overdose have risen steadily over the past two decades and have become the leading cause of injury death in the United States.(5) Since 1999, prescription opioid use and overdose deaths have quadrupled.(6) More than 165,000 people have died from prescription opioids in this timeframe,(7) yet there has not been an overall change in the amount of pain that Americans report.(8,9) In 2014, more than 14,000 people died from prescription opioid overdose, more than any year on record.(7) Higher opioid dosages are associated with increased risks for motor vehicle injury, opioid use disorder, and overdose.(2) The risk for overdose increases in a dose-dependent manner and lower dosages of opioids reduce the risk for overdose.(2) Improved opioid prescribing is an essential component of efforts to reduce opioid exposure, and ultimately risk of overdose.(2)

1c.4. Citations for data demonstrating high priority provided in 1a.3
1. S.524 - Comprehensive Addiction and Recovery Act of 2016. Available at: https://www.congress.gov/bill/114th-congress/senate-bill/524/text.
2. Dowell D, Haegerich TM, Chou R. CDC Guideline for Prescribing Opioids for Chronic Pain - United States, 2016. MMWR Recomm Rep. 2016;65:1-49. doi:10.15585/mmwr.rr6501e1. (PMID: 26987082).
3. HHS. ASPE Issue Brief: Opioid Abuse in the U.S. and HHS Actions to Address Opioid-Drug Related Overdoses and Deaths; 2015.

Available at: http://aspe.hhs.gov/basic-report/opioid-abuse-us-and-hhs-actions-address-opioid-drug-related-overdoses-and-deaths.
4. US Department of Health and Human Services. National Action Plan for Adverse Drug Event Prevention. Washington, DC; 2014. Available at: http://health.gov/hcq/ade.asp.
5. CDC. Web-based Injury Statistics Query and Reporting System (WISQARS). 2014. Available at: http://www.cdc.gov/injury/wisqars/fatal.html.
6. Chen LH, Hedegaard H, Warner M. Drug-poisoning Deaths Involving Opioid Analgesics: United States, 1999-2011. NCHS Data Brief. 2014;(166):1-8
7. CDC. Wide-ranging online data for epidemiologic research (WONDER). Atlanta, GA: CDC, National Center for Health Statistics; 2016. Available at: http://wonder.cdc.gov.
8. Chang H, Daubresse M, Kruszewski S, et al. Prevalence and treatment of pain in emergency departments in the United States, 2000 – 2010. Amer J of Emergency Med 2014; 32(5): 421-31.
9. Daubresse M, Chang H, Yu Y, Viswanathan S, et al. Ambulatory diagnosis and treatment of nonmalignant pain in the United States, 2000 – 2010. Medical Care 2013; 51(10): 870-878.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
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 subcriteria 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):*
Behavioral Health : Alcohol, Substance Use/Abuse, Mental Health : Alcohol, Substance Use/Abuse

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Overuse, Safety : Medication Safety

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.)*
http://pqaalliance.org/measures/default.asp

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

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** Cancer_Exclusion_RxHCC-_ICD-9_and_10_Codes.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
N/A

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Any member in the denominator with opioid prescription claims where the MED is greater than 120mg for 90 consecutive days or

longer*

*MED calculation is included in S.6 Numerator Details

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
The measurement year.

S.6. **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, 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.
Any member in the denominator with opioid prescription claims greater than 120mg MED for 90 consecutive days or longer* (See Table Opioids-A: Opioid Medications)

*Identifying members with prescription opioids that exceeded the MED threshold:
To identify members with prescription opioids that exceeded the MED threshold, each claim is to be converted into the MED using the appropriate conversion factor associated with the opioid product of that prescription claim (see Appendix A). The MED for each day's claims then are summed to determine the total MED for that day.
For each member in the denominator:
1. Calculate the MED for each opioid prescription claim during the measurement period, using the following equations:
• # of Opioid Dosage Units per day = (Opioid claim quantity) / (Opioid claim days supply)
• MED Daily Dose per claim = (# of opioid dosage units per day) X (# mg opioid per dosage unit) X (MED conversion factor)
2. Sum the daily MEDs of all opioid claims for each day to arrive at a total daily MED for each member.
3. Identify the days where the MED threshold is exceeded.
4. Any member, for whom the MED threshold is exceeded for 90 consecutive days or longer, meets the criteria for the MED component of the numerator.

Table Opioid-A: Opioid Medications (MED conversion factor)
buprenorphine patch (12.6) buprenorphine tab or film (10) butorphanol (7) codeine (0.15) dihydrocodeine (0.25)
fentanyl buccal or SL tablets, or lozenze/troche (0.13) fentanyl film or oral spray (0.18) fentanyl nasal spray (0.16)
fentanyl patch (7.2) hydrocodone (1) hydromorphone (4) levorphanol (11) meperidine (0.1)
methadone (3) morphine (1) opium (1) oxycodone (1.5) oxymorphone (3)
pentazocine (0.37) tapentadol (0.4) tramadol (0.1)

*Note: Injectables and Opioid cough and cold products and combination products containing buprenorphine and naloxone (e.g., BunavailTM, Suboxone®, Zubsolv®) are excluded from the MED calculations. Ionsys® (fentanyl transdermal patch) is also excluded as it is only for inpatient use; It is also only available through a restricted program under a Risk Evaluation and Mitigation Strategy (REMS)

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Any member with two or more prescription claims for opioids filled on at least two separate days, for which the sum of the days supply is greater than or equal to 15.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, Populations at Risk : Dual eligible beneficiaries

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
Any member with two or more prescription claims for opioids filled on at least two separate days, for which the sum of the days supply is greater than or equal to 15.

Table Opioid-A: Opioid Medications

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

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Any member with a diagnosis for Cancer or a Prescription Drug Hierarchical Condition Category (RxHCC) 8, 9, 10, or 11 for Payment Year 2015; or RxHCC 15, 16, 17, 18, or 19 for Payment Year 2016 (see list in S.11 and S.2b); or a hospice indicator (Medicare Part D) from the enrollment database.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Hospice exclusion: Exclude those members identified in the Medicare Enrollment Database as being enrolled in hospice.

Cancer exclusion: For Payment Year 2015: RxHCC 8, 9, 10, or 11. For Payment Year 2016: RxHCC 15, 16, 17, 18, or 19
ICD 9 and 10 Codes to Identify Cancer: Please see attachment in S2.b

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
The measure is stratified by the following lines of business for the health plan:

- Commercial
- Medicare
- Medicaid

Medicare Plans are further stratified by Low Income Subsidy status
Definition: Medicare Low Income Subsidy (LIS) - A subsidy paid by the Federal government to the drug plan for Medicare beneficiaries who need extra help with their prescription drug costs due to limited income and resources. Medicare beneficiaries apply for the LIS with the Social Security Administration or their State Medicaid agency.

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

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
N/A

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
N/A

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **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.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
Step One:
Calculate the denominator by identifying the number of all eligible members with two or more prescription claims for opioids filled on at least two separate days, for which the sum of the days supply is greater than or equal to 15.

Step Two:
Calculate the numerator by:
For each member in the denominator:
a. Calculate the MED for each opioid prescription claim during the measurement period, using the following equations:
• # of Opioid Dosage Units per day = (Opioid claim quantity) / (Opioid claim days supply)
• MED Daily Dose per claim = (# of opioid dosage units per day) X (# mg opioid per dosage unit) X (MED conversion factor)

b. Sum the daily MEDs of all opioid claims for each day to arrive at a total daily MED for each member.

c. Identify the days where the MED threshold is exceeded.

d. Any member, for whom the MED threshold is exceeded for 90 consecutive days or longer, meets the criteria for the MED component of the numerator.

Step Three:
Divide the number of members that met the criteria in numerator (Step Two d.) by the denominator (Step One) and multiply times 1000. The rate is reported as a proportion: XX out of 1,000 members.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
N/A

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
N/A

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Health Plan Medical and Pharmacy Claims. Health Plan member enrollment information.

S.25. **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.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Health Plan, Population : National, Population : State

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Other, Pharmacy
If other: The level of analysis for this measure is the prescription drug health plan, but it contains claims data from multiple care settings, including ambulatory, skilled nursing facility, pharmacy etc.

S.28. **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

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
PQA_High_Dose_testing_attachment-635986122942715331.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Use of Opioids at High Dosage in Persons Without Cancer
Date of Submission: 5/13/2016
Type of Measure:

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

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]
AND
If patient preference (e.g., informed decision making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):
- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR
- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;
OR
there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes
10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).
11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.
12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.
14. Risk factors that influence outcomes should not be specified as exclusions
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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.23*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ administrative claims	☒ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

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

The measure was tested in three different health plan data sources – the Medicare population, one commercial heath plan, and the Medicaid population.

For the Medicare population, data used for testing came from three different sources. The Medicare Part D Prescription Drug Event (PDE) claims were used for the identification of prescription drugs and cancer exclusions. To identify dates of birth and continuous enrollment, the Common Medicare Environment (CME) data source was used.

The data source for the Commercial population came from the health plans' enrollment data, medical claims, and prescription claims.

For the Medicaid population, the data used for testing came from Medicaid administrative claims. Six Medicaid plans covering four states were included in the testing using data from a Pharmacy Benefits Manager (PBM) organization. In addition, two other state-based plans were included in the testing using their state Medicaid administrative claims database. Medical claims were used to identify the cancer diagnoses, and the pharmacy claims were used for the identification of prescription drugs.

1.3. What are the dates of the data used in testing? Click here to enter date range

The testing from the Medicare and Commercial populations used administrative claims data from January 1st 2013 to December 31st 2013. The majority of testing used Medicaid prescription claims data from January 1st 2015-Decemer 31st 2015. The data from this time period were the most complete recent data available at the time of testing. Testing also included prescription claims data from one state's Medicaid plan from July 1st 2014-June 30th 2015.

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.26*)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☒ health plan	☒ health plan
☐ other: Click here to describe	☐ other: Click here to describe

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

For the Medicare testing, the analysis included a convenience sample of over 700 Medicare Part D prescription drug plans.

Testing was also conducted in one Commercial health plan. The size and characteristics of these populations are included at the patient level in 1.6.

For the Medicaid testing, the analysis included 8 state based prescription drug plans covering 6 states. 3 plans were from the same state in the Mid-Atlantic region of the United States (US), 2 plans were from states in the South Atlantic region of the US, two plans were from states in the West South Central region of the US, and one plan was from a state in the East South Central region of the US. The size and characteristics of the population are included at the patient level in 1.6.

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

For the Medicare testing, a total of 7,067,445 individuals aged 18 and older were included in the testing and analysis. This data can be stratified by age, gender, and type of Part D plan. Of all persons, 2,531,712 (35.8%) are male, and 4,535,732 (64.2%) are female. Individuals by age group included 271,635 (3.8%) age 18-40, 2,159,384 (30.6%) age 41-64 and 4,636,425 (65.6%) over age 65. Of all individuals, 2,492,658 (35.3%) are enrolled in a Medicare Advantage Prescription Drug Plan (MA-PD) and 4,574,787 (64.7%) are enrolled in a standalone Prescription Drug Plan (PDP).

For the Commercial plan, a total of 209,191 individuals age 18 and older were included in the analysis. Of all persons 92,227 (44.1%) are male, and 116,964 (55.9%) are female. Persons by age group included 46,913 (22.4%) age 18-40, 133,207 (63.7%) age 40-64 years, and 29,071 (13.9%) age 65 and older.

For the Medicaid plans, a total of 1,437,410 individuals age 18 and older were included in the analysis. Of all persons 515,164 (35.8%) are male, and 922,246 (64.2%) are female. Persons by age group included 897,641 (62.4%) age 18-40, 454,528 (31.6%) age 40-64 years, and 85,241 (6.0%) age 65 and older.

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

The Medicaid data was used to test reliability, as this measure has been added to the Medicaid Adult Core Set. This data does not include the RxHCC indicator to identify cancer exclusions, and instead uses ICD-9 or ICD-10 (depending on the year of the data) to identify diagnostic criteria for the cancer exclusions. For the majority of the plans, the Medicaid data also does not allow for identification of hospice patients.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

For the Medicare population, the beneficiary level Low Income Subsidy (LIS) variable was used to determine disparities in rates for populations with different sociodemographic status. The LIS is a subsidy paid by the Federal government to the drug plan for Medicare beneficiaries who need extra help with their prescription drug costs due to limited income and resources. For the Commercial and Medicaid other populations, no patient level indicators of sociodemographic status were available in the data.

2a2. RELIABILITY TESTING

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

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

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

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

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

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

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

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

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

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

Using the parameter estimates from the Beta-Binomial model we computed individual plan reliability scores. Table 1 below shows the distribution of the plan-level scores. Plans have very high reliability scores. The reliability score mean is 0.9938 and the median 0.9945.

Table 1. Individual Plan Reliability Score Distribution

Statistic	Values
Mean	0.9938
Standard Dev.	0.0044
Min	0.9843
p10	0.9895
p25	0.9934
p50 (Median)	0.9945
p75	0.9958
p90	0.9971
max	0.9995

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

The reliability score mean is 0.9938 and the median 0.9945. A reliability score of 0.7 is the minimum threshold for reliability. Based on the high reliability scores for each of the plans in the analysis, the measure is considered reliable.

2b2. VALIDITY TESTING

2b2.1. 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*)

2b2.2. 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)*

PQA uses a systematic, transparent, consensus-based measure development and testing process. That process used in 2014 to develop this measure is outlined below:

- Step 1: PQA workgroups identify measure concepts that may be appropriate for development into fully specified performance measures. The workgroups focus on specific aspects of the medication-use system and/or specific therapeutic areas. The workgroups are open to all members of PQA and use a consensus-based approach to identify, prioritize and recommend the measure concepts that are deemed to be highly important for supporting quality improvement related to medications.

- Step 2: The measure concepts that are recommended for further development through a vote by the PQA workgroups are forwarded to the PQA Quality Metrics Expert Panel (QMEP) for evaluation and refinement. The QMEP is composed of PQA members who have backgrounds in pharmacy, medicine, research, quality improvement and measures development. The QMEP reviews the measure concepts to provide an initial assessment of the key properties of performance measures (i.e., feasibility, usability and scientific validity). The measure concepts that are rated highly on these key properties will then undergo technical specification.

- Step 3: The draft measure is provided to PQA member organizations for their comments prior to preparing technical specifications for pilot testing. The QMEP reviews member comments, edits the draft measure accordingly and poses testing questions based on this all-member feedback.

- Step 4: PQA selects partners to test the draft measure. These partners are often PQA member health plans or academic institutions with expertise in quality and performance measure testing. The testing partner implements the draft technical specifications with their existing datasets and provides a report to PQA that details testing results and recommendations for modifications of the technical specifications.

- Step 5: The workgroup that developed the measure reviews the testing results and provides comment. The QMEP reviews the workgroup comments, testing results, recommendations and potential modifications and provides a final assessment of the feasibility and scientific validity of the draft performance measures.

- Step 6: Measures that are recommended by the QMEP for endorsement are posted on the PQA web site for member review, written comments are requested, and a conference call for member organizations is scheduled to address any questions. This process allows members to discuss their views on the measures in advance of the voting period.

- Step 7: PQA member organizations, which include organizations such as large pharmacy chains, health plans, quality organizations and pharmaceutical companies vote on the performance measure(s) considered for approval and/or endorsement.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

The measure was assessed for face validity (i.e., whether it appears to measure what it intends to measure) through

review by the PQA workgroup that developed the measure, the PQA Quality Metrics Expert Panel (QMEP), and PQA's full membership. In addition, feedback about validity of the measure was sought out by the five PQA member organizations who tested the measure using their own data.

The PQA Medication Use Safety Workgroup was composed of 72 PQA members that worked on multiple measure concepts. After the workgroup completed the development of the measure specifications, 37 members of the workgroup voted to determine if the draft measure should continue on further development and review by the PQA QMEP. 94.6% of members recommended that the measure move on for QMEP review.

The PQA QMEP is a panel that includes individuals with expertise and experience in pharmacy, medicine, research, and clinical or other technical expertise related to quality improvement and measure development. The names and credentials of the QMEP Panel are listed in Table 1. The QMEP reviewed the measure prior to testing to ensure scientific soundness and usefulness. The QMEP reviewed the results of the measure testing including the performance measure scores reported by plan referenced in Section 2b5 (below). Out of the 12 members of the QMEP who voted, 67% strongly agreed that the measure results reflected the quality of care, and recommended that the measure be considered for endorsement by the PQA membership.

Table 1. **PQA Quality Metrics Expert Panel (QMEP)**

QMEP Member Name and Credentials	QMEP Member Organization
Bimal Patel, Pharm D, MS	MedImpact
Catherine Coast, PharmD	Highmark
Chris DuPaul, MBA	CVS Caremark
Christopher Dezii, RN, MBA, CPHQ	Bristol-Myers Squibb
Christopher Powers, PharmD	CMS
David Nau, PhD, RPh, CPHQ	Pharmacy Quality Solutions
Gary Erwin, PharmD	OmniCare
Gary Young, JD, PhD	Northeastern University
Jenny Weber, PharmD, MS, PCPS,CGP, BCACP	Humana
Jessica Frank, PharmD	OutcomesMTM
Karen Farris, PhD	University of Michigan
Keith Widmer, RPh, BCPP	Express Scripts
Kent Summers, RPh, PhD	Astellas
Lynn Deguzman, PharmD, CGP	Kaiser Permanente
Mary Ann Kliethermes, PharmD	Midwestern University
Mitzi Wasik, PharmD, PCPS	Coventry Health Care/Aetna
Pat Gleason, Pharm D, BCPS	Prime Therapeutics
Steve Riddle, PharmD, BCPS	Wolters Kluwer Health
Steven Burch, RPh, PhD	GlaxoSmithKline
Tony Willoughby, PharmD	HealthMart-McKesson

PQA membership was notified prior to the PQA Annual Meeting in May 2015, of the opportunity to consider and vote for the performance measure during the meeting. (Note: PQA membership comprises health plans, community pharmacy, long-term care pharmacies, HIT companies, PBMs, healthcare quality and standards organizations, professional and trade associations, and others.) Members received the measure description, key points and evidence, measure specifications, and the performance measure scores reported by plan. During the PQA Business meeting, the measure was reviewed. Nearly all of PQA membership had a representative at the Annual Meeting and were present for the vote. Voting options included, "Agree" (indicating that the organization approved the measure), "Disagree (indicating that the organization opposed the measure) and "Abstain." Out of the 89 number of PQA members who participated in voting, 69.7% of the membership voted in favor of endorsing the measure.

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

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

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

2b3. EXCLUSIONS ANALYSIS
NA ☐ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

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

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

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

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

2b3.2. 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*)

Of the eight health plans included in the analysis, the cancer patient exclusions were 0.5% to 1.9% of the overall population. The one Medicaid plan that could identify hospice exclusions found only 15 cases, which represented 0.003% of their total population.

2b3.3. 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 results show that in some plans, almost 2% of the population has cancer and would be included in the measure if cancer was not excluded. This is a significant proportion of the population that could potentially impact the measure rates. No inferences about the hospice exclusion could be drawn because the majority of the plans could not identify exclusions.

2b4. 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 2b5.

2b4.1. What method of controlling for differences in case mix is used?
☐ **No risk adjustment or stratification**
☐ **Statistical risk model with** Click here to enter number of factors **risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use 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.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

2b4.4a. What were the statistical results of the analyses used to select risk factors?

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

2b4.5. 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 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

2b4.10. 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*)

2b4.11. 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*)

_____

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

To assess significant differences in measure rates, the data described in sections 1.5 and 1.6 above were used to calculate the mean, median, standard deviation, and interquartile range for the measure rates for the Medicare population and the Medicaid population. In addition, for the Medicaid population, the rates were divided into quartiles, and a Student's t-test was used to compare the rates between the bottom quartile (75th percentile) and top quartile (25th percentile). A student's t-test was used to compare the rates of the plans in the 25th percentile to the plans with rates in the 75th percentile. The statistics are for the Medicare population is reported below in 2b5.2, Tables 1 and 2. The statistics for the Medicaid population is reported below in 2b5.2, Tables 3 and 4.

2b5.2. 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 1. Variation in Measure Rates - Medicare Population (reported as number per 1,000 members)

Mean	Median	Standard Deviation
39.27	38.70	8.32

Table 2. Interquartile Range of Measure Rates - Medicare Population (reported as number per 1,000 members)

Minimum	30.00
25th Percentile	34.62
50th Percentile	38.70
75th Percentile	43.35
Maximum	49.66
Interquartile Range	8.73

Table 3. Variation in Measure Rates - Medicaid Population (reported as number per 1,000 members)

Mean	Median	Standard Deviation
34.04	34.29	20.61

Table 4. Interquartile Range of Measure Rates - Medicaid Population (reported as number per 1,000 members)

Minimum	8.15
25th Percentile	20.4
50th Percentile	34.3
75th Percentile	48.1
Maximum	66.45
Interquartile Range	27.68
Student's t-test p-value	0.029

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i*.e., what do the results mean in terms of statistical and meaningful differences?*)

For the Medicare population, the measure rates showed significant variation, with a standard deviation of 8.32 and an Interquartile Range was 8.73.

For the Medicaid population, the measure rates showed even greater variation, with a standard deviation of 20.61 and an Interquartile Range of 27.68.

There is a statistically significant difference in measure rates between the top and bottom quartile of the plans included in the testing (*P*=0.029 at alpha=0.05). This variation shows that there are meaningful differences in rates across plans.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
If only one set of specifications, this section can be skipped.

Only one set of specifications is provided for this measure.

Note: *This item is directed to measures that are risk-adjusted (with or without SDS 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 SDS 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.***

2b6.1. 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*)

2b6.2. 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*)

2b6.3. 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*)

_____

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

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

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

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

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

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

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

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

3. Feasibility
Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.
3a. Byproduct of Care Processes

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

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), Other
If other: Prescription claims data

3b. Electronic Sources

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

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*
ALL data elements are in defined fields in electronic claims

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

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.**
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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
Pilot test sites indicated the measure was feasible and results were able to be reported efficiently and accurately. CMS calculates the measure for Part D plans. The data is readily available (prescription claims data and medical data).

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
PQA develops and maintains numerous performance measures related to the medication use system. The Measures are the proprietary property of PQA, and it is in the interest of PQA to protect and promote the appropriate use of the Measures. PQA may approve an organization´s use of the Measures; however, no organization may use the Measures without first obtaining permission from PQA prior to using the Measures. Certain uses of the Measures are only approved with a licensing agreement from PQA that specifies the terms of use and the licensing fee. PQA reserves the right to determine the conditions under which it will approve and/or license the Measures.

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. **Current and Planned Use**
NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within

6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
Public Reporting Quality Improvement (Internal to the specific organization)	Quality Improvement with Benchmarking (external benchmarking to multiple organizations) CMS Medicare Part D - Patient Safety Reports http://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovGenIn/index.html

4a.1. **For each CURRENT use, checked above, provide:**

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

Name of program and sponsor: CMS Medicare Part D Drug Benefit
Purpose: Monitor Opioid use by Medicare Part D beneficiaries
Geographic area: National, approximately 38 million beneficiaries in Medicare Part D plans.

4a.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 was developed in 2015.

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
The measure is currently being used in the Medicare Part D Overutilization Monitoring System to monitor the utilization of opioids for members with the Medicare drug benefit.

CMS has announced plans to move this measure into the 2019 Part D Display Measures, using data from 2017.

This measure also has been included in the 2016 Medicaid Adult Core Set.

Reporting of results is not yet available.

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

N/A - initial endorsement submission.

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
N/A

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**

No unintended negative consequences to individuals or populations were identified during testing. This measure, Use of Opioids at High Dosage in Persons Without Cancer, has been implemented by CMS Part D as part of the Overutilization Monitoring System beginning January, 2016. To date, no negative consequences have been identified.

However, concerns have been raised that prescribing changes such as dose reduction (without offering or arranging evidence-based treatment for patients with opioid use disorder) might be associated with unintended negative consequences, such as patients seeking heroin or other illicitly obtained opioids (1,2) or interference with appropriate pain treatment.(3) Data indicate that if access to prescription opioids is limited, some users of opioid analgesics will transition to heroin or other illicitly obtained opioids, leading to increased overdose death coincident with prescribing restrictions.(1) There are also concerns about pain being underdiagnosed and undertreated, particularly for ethnic and racial minorities, females, children or infirm elderly, or individuals who may be perceived to have mental health problems or are drug seeking.(4,5)

These concerns must be balanced by the current situation in the United States which has been described by the CDC as an epidemic of opioid abuse, overdose, and deaths. Overdose involving opioid analgesics killed almost 17,000 persons in 2010 and the number of people with opioid analgesic use disorders increased to nearly 2 million.(6) We believe the potential benefits of monitoring those patients receiving the very highest doses of opioids for extended periods of time or receiving these drugs from multiple providers outweighs potential negative consequences.

References:

1. Coffin P, Banta-Green C. The dueling obligations of opioid stewardship. Ann Intern Med. 2014;160:207–8. doi.org/10.7326/ M13-2781. (PMID 25133372).
2. Cicero,T, Ellis M, Harney J. Shifting Patterns of Prescription Opioid and Heroin Abuse in the United States. N Engl J Med. 2015; 373:1789-90. DOI: 10.1056/NEJMc1505541. (PMID 26510045).
3. Twillman RK, Kirch R, Gilson A. Efforts to control prescription drug abuse: Why clinicians should be concerned and take action as essential advocates for rational policy. CA Cancer J Clin. 2014;64:369–76. doi.org/10.3322/caac.21243. (PMID 25044063).
4. Kirschner N, Ginsburg J, Snyder LS, Health and Public Policy Committee of the American College of Physicians. Prescription Drug Abuse; executive summary of a policy position paper from the American College of Physicians. Ann Intern Med. 2014;160:198-200. doi:10.7326/M13-2209. (PMID 24323199).
5. Institute of Medicine (US) Committee on Advancing Pain Research, Care, and Education. Relieving Pain in America: A Blueprint for Transforming Prevention,Care, Education, and Research. Washington (DC): National Academies Press (US); 2011. (PMID: 22553896). Available at: http://www.nap.edu/read/13172/chapter/1
6. Coffin P, Banta-Green C. The dueling obligations of opioid stewardship. Ann Intern Med. 2014;160:207. doi: 10.7326/M13-2781. (PMID 24322334).

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

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 completely harmonized?

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.)
N/A

Appendix

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

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** PQA
Co.2 **Point of Contact:** Julie, Kuhle, jkuhle@pqaalliance.org, 515-554-6685-
Co.3 **Measure Developer if different from Measure Steward:** PQA
Co.4 **Point of Contact:** Julie, Kuhle, jkuhle@pqaalliance.org, 515-554-6685-

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

PQA Quality Metrics Expert Panel 2015
Responsible for review and consideration of the measure concept and all testing results of the draft measure

Bimal Patel*, Pharm D, MS	MedImpact
Catherine Coast, PharmD	Highmark
Chris DuPaul, MBA	CVS Caremark
Christopher Dezii,RN, MBA, CPHQ	Bristol-Myers Squibb
Christopher Powers, PharmD	CMS
David Nau, PhD, RPh, CPHQ	PQS
Gary Erwin, PharmD	Omnicare
Gary Young, JD, PhD	Northeastern University
Jenny Weber*, PharmD, MS, PCPS,CGP, BCACP	Humana
Jessica Frank, PharmD	OutcomesMTM
Karen Farris, PhD	University of Michigan
Keith Widmer, RPh, BCPP	Express Scripts

Kent Summers, RPh, PhD	Astellas
Lynn Deguzman, PharmD, CGP	Kaiser Permanente
Mary Ann Kliethermes, PharmD	Midwestern University
Mitzi Wasik, PharmD, PCPS	Coventry Health Care/Aetna
Pat Gleason, Pharm D, BCPS	Prime Therapeutics
Steve Riddle, PharmD, BCPS	Wolters Kluwer Health
Steven Burch, RPh, PhD	GlaxoSmithKline
Tony Willoughby, PharmD	HealthMart-McKesson

* denotes co-chair

PQA Medication Use Safety Workgroup 2014

Responsible for development of the measure

Amber Baybayan OutcomesMTM
David Belew MedHere Today
Rachael Boggs PQA Invited Guest Participant
Stay Bontha PerformRx
Sara Burnheimer UPMC Health Plan
Patrick Campbell University of Arizona College of Pharmacy
Scott Campbell PQA Invited Guest Participant
Rebecca Chater Ateb
Trina Clark GlaxoSmithKline
Victor Cohen American Society of Health-System Pharmacists (ASHP)
Michael Contos Indian Health Services
Karen Davidson Therapeutic Research Center (home of Pharmacist's Letter and Prescriber's Letter)
Shelly Delaville American Society of Consultant Pharmacists (ASCP)
James DeVita CVS/Caremark
Sara Ericsson MedImpact Healthcare Systems, Inc.
Marybeth Farquhar URAC
Alison Farrell Ahold USA
Cindi Fitzpatrick U.S. Food & Drug Administration (FDA)
Jeremy Fredell Express Scripts, Inc.
George Garmer CARE Pharmacies Cooperative
Jennifer Gatsos-Walter Wolters Kluwer Health, Clinical Solutions
Mary Ghods U.S. Food & Drug Administration (FDA)
James Glass Rite Aid
Averill Gordon Walgreen Co.
Lindsey Gumbo Pharmaceutical Research & Manufacturers of America (PhRMA)
Tracy Harrell SinfoníaRx
Tiffany Harris SCAN Health Plan
Shannon Harrison Highmark Health Services
Lisa Hines* University of Arizona College of Pharmacy
John Kessler National Alliance of State Pharmacy Associations (NASPA)
Mi´a Kirkland Wellcare
Nicholas Kostek Kaiser Permanente
Maribeth Kowalski Purdue Pharma, L.P.
Jason Kinsman RxAnte
Edward Lennard U.S. Office of Personnel Management
Patricia Marchlowska Lilly USA
Peter Marshall HealthPartners
Kevin Masci Target
Richard McLeod Pfizer, Inc.
Diane McNally Centers for Medicare & Medicaid Services (CMS)
Brent Merrick Cigna-HealthSpring
Leslie Miller Gorman Health Group
Joel Montavon Catamaran
Kim Moon PQA Invited Guest Participant
Gina Moore PQA Invited Guest Participant

Scott Nakagawa Applied Research Works
Patricia Neafsey ActualMeds Corporation
Jeffrey Nesheim Takeda Pharmaceuticals America, Inc.
Michael Nguyen CenseoHealth
Kyle Null* University of Mississippi Center for Pharmaceutical Marketing & Management
Udo Nwachukwu Mirixa Corporation
Steven Oh Health Mart Systems Inc.
Maria Osborne American Pharmacists Association (APhA)
Nicole Paterson Fairview Medication Therapy Management
Jacqui Pesa Johnson & Johnson
Roger Pinsonneault RelayHealth
Richard Segal University of Florida College of Pharmacy
Bupendra Shah Long Island University Arnold & Marie Schwartz College of Pharmacy
Christine Sommer First DataBank
Catherine Starner Prime Therapeutics
Karen Stockl UnitedHealth Group
Brian Sweet AstraZeneca
Christie Teigland Inovalon, Inc.
Jennifer Thomas National Alliance of State Pharmacy Associations (NASPA)
Ly Tran PharmMD
Maria Vassilakis Astellas Scientific and Medical Affairs, Inc.
Kathleen Vest PQA Invited Guest Participant
Brandi Rosberg Walmart
Jennifer Weber Humana
Elizabeth Whaley-Buono MeadWestvaco
Jennifer Williams Aetna
Melissa Wilson Capital Health Plan

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2015

Ad.3 **Month and Year of most recent revision:** 05, 2015

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

Ad.5 **When is the next scheduled review/update for this measure?** 10, 2016

Ad.6 **Copyright statement:** Rights Retained by PQA, Inc 2016.

Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 2950
De.2. **Measure Title:** Use of Opioids from Multiple Providers in Persons Without Cancer
Co.1.1. **Measure Steward:** PQA
De.3. **Brief Description of Measure:** The proportion (XX out of 1,000) of individuals without cancer receiving prescriptions for opioids from four (4) or more prescribers AND four (4) or more pharmacies.
1b.1. **Developer Rationale:** Abuse and overdose of prescription drugs is a major public health issue in the United States.(1,2) Studies have shown that people who see multiple prescribers or use multiple pharmacies are more likely to die of drug overdoses.(3)

PQA developed 3 measures related to prescription opioid use that are indicative of the quality of care for patients taking these medications. The measures examine the quality of use related to the dose of the medications over time, access to the medications through multiple providers, and the combination of both these criteria. This measure, Use of Opioids from Multiple Providers in Persons Without Cancer, focuses specifically on the use of opioids from multiple providers.

Prescription drug monitoring programs (PDMP), which track the use of multiple providers by patients, indicate that such use is typically found among a small proportion of patients, with the proportion declining as the number of providers increases. In Massachusetts in 2006, considering only Schedule II opioids, 0.5% of patients saw 4+ prescribers and 4+ pharmacies.(4) A national study found that 13% of patients had overlapping prescriptions from two or more different prescribers during an 18-month period. Of these, 0.5% used 4+ prescribers and 4+ pharmacies.(5) People who see multiple prescribers or use multiple pharmacies are more likely to die of drug overdoses.(3) When comparing the diagnostic odds ratio for opioid overdose events of 9 pharmacy shopping definitions, a threshold of 4 pharmacies had the highest diagnostic odds ratio.(6) Data from the California PDMP indicates that people with higher daily dosages are more likely to see multiple prescribers or go to multiple pharmacies.(7) However, there is no clear threshold at which multiple prescribers and multiple pharmacies represent lack of continuity or poorly coordinated care.

Data suggest that efforts to prevent opioid overdose deaths should include a multi-faceted approach focused on strategies that target high-dose opioid users as well as persons who seek care from multiple doctors and pharmacies. The data also suggests that these criteria can be considered separately, as measures related to prescribed opioids for appropriate clinical uses versus inappropriate uses. Thus, as stated above, PQA developed 3 measures: one for high dose therapy, one for multiple providers, and one that is the intersection of both high dose and multiple providers – with this measure presently under consideration focused specifically on the use of opioids from multiple providers.

References:
1. U.S. Department of Health and Human Services, Office of Disease Prevention and Health Promotion. (2014). National Action Plan for Adverse Drug Event Prevention. Washington, DC. Accessed on: 4/9/15. Available at: http://www.health.gov/hcq/pdfs/ADE-Action-Plan-508c.pdf.
2. Liu Y, Logan JE, Paulozzi LJ, et al. Potential misuse and inappropriate prescription practices involving opioid analgesics. Am J Manag Care. 2013;19:648-65. PMID: 24304213.
3. Paulozzi LJ, Kilbourne EM, Shah NG, et al. A history of being prescribed controlled substances and risk of drug overdose death. Pain Med. 2012;13:87-95. PMID: 22026451.
4. Katz N, Panas L, Kim M, et al. Usefulness of prescription monitoring programs for surveillance--analysis of Schedule II opioid prescription data in Massachusetts, 1996- 2006. Pharmacoepidemiol Drug Saf. 2010;19:115-23. PMID: 20014166.
5. Cepeda MS, Fife D, Chow W, et al. Assessing opioid shopping behaviour: a large cohort study from a medication dispensing database in the US. Drug Saf. 2012;35:325-34. PMID: 22339505.
6. Yang Z, Wilsey B, Bohm Michele, et. al. Defining Risk of Prescription Opioid Overdose: Pharmacy Shopping and Overlapping Prescriptions Among Long-term Opioid Users in Medicaid. The Journal of Pain. 2015;445—453. PMID 25681095.

7. Han H, Kass PH, Wilsey BL, et al. Individual and county-level factors associated with use of multiple prescribers and multiple pharmacies to obtain opioid prescriptions in California. PLoS One. 2012;7:e46246. PMID: 23049992.

S.4. **Numerator Statement:** Any member in the denominator who received opioid prescription claims from 4 or more prescribers AND 4 or more pharmacies.
S.7. **Denominator Statement:** Any member with two or more prescription claims for opioids filled on at least two separate days, for which the sum of the days supply is greater than or equal to 15.
S.10. **Denominator Exclusions:** Any member with a diagnosis for Cancer or a Prescription Drug Hierarchical Condition Category (RxHCC) 8, 9, 10, or 11 for Payment Year 2015; or RxHCC 15, 16, 17, 18, or 19 for Payment Year 2016; (see list in S.11 and S.2b); or a hospice indicator from the enrollment database.

De.1. **Measure Type:** Process
S.23. **Data Source:** Administrative claims
S.26. **Level of Analysis:** Health Plan, Population : National, Population : State

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

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *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.

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 evidence suggests that prescriptions for opioids from multiple prescribers and pharmacies correlates with undesired health outcomes. The use of multiple prescribers and pharmacies are associated with increased risks for opioid overdose.

Guidance from the Evidence Algorithm
1-No→3-No→7-Yes→ 8-Yes→MODERATE

Questions for the Committee:

- *What is the relationship of this measure to patient outcomes?*
- *How strong is the evidence for this relationship?*
- *Is the evidence directly applicable to the process of care being measured?*

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

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

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

- The measure was tested in three different health plan data sources – the Medicare population, one commercial heath plan, and the Medicaid population.
- The testing from the Medicare population used administrative claims data from January 1st 2013 to December 31st 2013. The Medicare rates ranged from 30.0 per 1,000 to 49.66 per 1,000. The Mean was 39.27 per 1,000

and the median was 38.7 per 1,000. The standard deviation was 8.32. The 25th percentile was 34.62 per 1,000, the 50th percentile is the median (38.70 per 1,000) and the 75th percentile was 43.35 per 1,000. The interquartile range was 8.73.

- The Medicaid rates ranged from 8.15 per 1,000 to 66.45 per 1,000. The Mean was 34.04 per 1,000 and the median was 34.29 per 1,000. The standard deviation was 20.61. The 25th percentile was 20.4 per 1,000, the 50th percentile is the median (34.29 per 1,000) and the 75th percentile was 48.1 per 1,000. The interquartile range was 27.68.
- Testing was also conducted in one Commercial health plan using administrative claims from January 1st 2013 to December 31st 2013. This plan covered 209,191 individuals age 18 and older. The measure rate for this plan was 32.03 per 1,000.

Disparities

- The beneficiary level Low Income Subsidy (LIS) variable was used to determine disparities in rates for populations with different sociodemographic status. The LIS is a subsidy paid by the Federal government to the drug plan for Medicare beneficiaries who need extra help with their prescription drug costs due to limited income and resources. The measure rate for the LIS group is 62.41 per 1,000 while the rate for the non-LIS population is significantly lower, at 28.09 per 1,000.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*

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

Comments: ** This is a process measure. There was no systematic review of the evidence specific to this measure. The evidence provided suggests that prescriptions for opioids from multiple prescribers and pharmacies is correlated with undesirable outcomes., including risk for opioid overdose. Low to moderate rating for evidence.\

**There is moderate evidence to suggest that opioids from multiple providers can lead to an increase risk to patients

1b. Performance Gap

Comments: **The measure was tested in 3 health plan data sources - Medicare, commercial and Medicaid - using administrative claims data The analysis showed a wide range of performance . The developer used the beneficiary low income subsidy variable to determine disparities in rates for populations with different SDS. There was a significantly lower rate for the non-LIS population than the LIS population.

**The data suggests disparity a difference in this rate between different SES groups

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Administrative claims, health plan enrollment information

Specifications:

- This measure assesses the proportion of individuals without cancer receiving prescriptions for opioids from four or more prescribers AND four or more pharmacies.
- The level of analysis (i.e., the measured entity) is the prescription drug health plan.
    - The developer notes that the measure also contains claims data from multiple care settings, including ambulatory, skilled nursing facility, pharmacy etc.
- The measure is stratified by the following lines of business for the health plan:
    - Commercial
    - Medicare
    - Medicaid
- The measure is reported as a rate (per 1,000 plan members).
- The measure uses health plan medical and pharmacy claims and health plan member enrollment information as its data sources.
- To identify the denominator population, the measure identifies any member with two or more prescription claims for opioids filled on at least two separate days, for which the sum of the days' supply is greater than or equal to 15.
- To derive the numerator, the measure calculates the number of unique pharmacy providers associated with an opioid prescription claim and the number of unique prescribers associated with an opioid prescription claim. Any member with four or more unique pharmacy providers AND four or more unique prescribers meets the criteria for the numerator.
- A list of opioid medications is provided in the submission form.
- The measure excludes patients with a diagnosis of cancer and patients in hospice.
- A list of administrative codes (ICD-9/10, RxHCC) identifying denominator exclusions is provided in a spreadsheet attached to the measure submission.

Questions for the Committee :

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

SUMMARY OF TESTING

Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

- The developer used several data sets for reliability testing:
    - For Medicare testing, the analysis included a convenience sample of over 700 Medicare Part D prescription drug plans (comprising a total of 7,067,445 individuals aged 18 and older)
    - Testing was also conducted in one Commercial health plan (comprising a total of 209,191 individuals age 18 and older)
    - For Medicaid testing, the analysis included 8 state-based prescription drug plans covering 6 states (comprising a total of 1,437,410 individuals age 18 and older)

Method(s) of reliability testing

- To demonstrate reliability, the developer conducted a signal-to-noise analysis of the computed measure score using a beta-binomial model.
    - The developer explains that a reliability score (i.e., signal-to-noise ratio) may range from 0 to 1; a score

of 0 signifies that all variation is due to measurement error ("noise"), while a score of 1 signifies that all variation represents true differences in performance scores between plans ("signal").

Results of reliability testing

- The developer provides the results of reliability testing in a table presenting the distribution of individual plan reliability scores; the mean reliability score across all plans is **0.9355**.
- The developer suggests that a reliability score of 0.7 is the minimum threshold for reliability, and that based on the high scores achieved in the analysis, this measure should be considered reliable.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

Guidance from the Reliability Algorithm

[Box 1] Specifications precise and unambiguous → [Box 2] Empirical testing conducted on the measure as specified → [Box 4] Testing conducted at the measure score level → [Box 5] → Testing method described and appropriate → [Box 6] High certainty or confidence that measure scores are reliable → [Box 6a]

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

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☒ **Somewhat** ☐ **No**

Specification not completely consistent with evidence

- The evidence cited by the developer is not entirely unanimous (though it is nearly so) about the number of unique prescribers that signals potential opioid abuse.

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

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.

SUMMARY OF TESTING

Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

- ☒ **Face validity only**
- ☐ **Empirical validity testing of the measure score**

Validity testing method:

- To demonstrate validity, the developer cites their (PQA's) approach to measure development and testing.
  - This approach includes identification of important concepts by PQA member workgroups, evaluation and refinement of concepts by the PQA Quality Metrics Expert Panel (QMEP), partnership with measure development experts, and processes for review, comment, and approval by PQA members.
- The developer notes that the QMEP Panel reviewed the results of measure testing, including performance

measure scores, and provided an assessment of whether measure results reflect quality of care.

Validity testing results:

- The developer reports that out of 12 QMEP members voting on the measure's face validity, 67 percent strongly agreed that the measure results reflected quality of care.
- In addition, the developer notes that of 89 PQA members voting on whether to endorse the measure, 69.7 voted in favor of approval.
- Five PQA member organizations also tested the measure using their own data, and all strongly agreed that the measure reflected the quality of care provided for their populations.

- ***NQF Staff Note:*** *Assessment of this measure's validity appears to have been conducted by the same groups involved in development of the measure; NQF prefers face validity to be assessed by experts or other stakeholder groups who have not been involved in the measure development process.*

Questions for the Committee:

- *Does the information provided by the developer demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- This measure excludes patients with a diagnosis of cancer and patients in hospice.
- The developer's rationale for these exclusions is that patients at end of life, undergoing hospice care, and those with cancer may have unusual requirements for pain management; the developer notes that these exclusions are consistent with the 2016 CDC Guideline for Prescribing Opioids for Chronic Pain.
- Because prescription claims data do not contain claims for palliative medication, such as opioids, for persons in Medicare Part D that are in hospice care, this exclusion was not tested by the developer. In addition, for the Medicaid population, the majority of the plans were not able to identify hospice exclusions in their data.
- For the cancer exclusion, the developer provided an analysis of data from eight health plans, identifying the number of exclusions and the percent of the overall population that would be affected by including patients with cancer diagnoses.
- The developer reports that the cancer patient exclusions were 0.5% to 1.9% of the overall population. The one Medicaid plan that could identify hospice exclusions found only 15 cases, which represented 0.003% of their total population.
- Interpreting the results of this analysis, the developer states that in some plans, almost 2% of the population has cancer and would be included in the measure if cancer was not excluded, suggesting that this is a significant proportion of the population that could potentially impact the measure rates.
- The developer states that no inferences about the hospice exclusion could be drawn because the majority of the plans could not identify exclusions.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

Questions for the Committee:

- *Do you agree with the developer that this measure does not require risk adjustment?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- To assess the measure's ability to identify meaningful differences in performance, the developer analyzed their testing data to identify the mean, median, standard deviation, and interquartile range for the measure rates for the Medicare population and the Medicaid population.
- In addition, for the Medicaid population, the rates were divided into quartiles, and a Student's t-test was used to compare the rates between the bottom quartile (75th percentile) and top quartile (25th percentile).
- For the Medicare population, the developer reports a mean performance rate (per 1,000 members) of **23.31**, a median rate of **26.12**, and a standard deviation of **5.73**.
- For the Medicaid population, the developer reports a mean performance rate (per 1,000 members) of **72.28**, a median rate of **69.93**, and a standard deviation of **12.03**.
- The following tables provide additional results of the developer's analysis:

Interquartile Range of Measure Rates - Medicare Population (reported as number per 1,000 members)

Minimum	17.98
25th Percentile	22.32
50th Percentile	23.12
75th Percentile	29.10
Maximum	31.00
Interquartile Range	6.78

Interquartile Range of Measure Rates - Medicaid Population (reported as number per 1,000 members)

Minimum	57.56
25th Percentile	65.13
50th Percentile	69.93
75th Percentile	80.66
Maximum	93.08
Interquartile Range	15.51
Student's t-test p-value	0.034

- The developer's interpretation of these results is that the measure rates showed significant variation in the Medicare population, and even greater variation in the Medicaid population
- The developer also states that there is a statistically significant difference in measure rates between the top and bottom quartile of the plans included in the testing (P=0.034 at alpha=0.05), and suggests that this variation shows that there are meaningful differences in rates across plans.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

N/A

2b7. Missing Data

- The developer notes that since all data elements are available via prescription claims data, it is not expected—nor was it found—that missing data would result.
- The developer states that, as a result, performance results would not be biased, as prescription claims data provides the data elements necessary to calculate the measure rate.

Guidance from the Validity Algorithm

[Box 1] Specifications consistent with evidence → [Box 2] Potential threats to validity addressed → [Box 3] Empirical validity testing **NOT** conducted using the measure as specified → [Box 4] Face validity systematically assessed → [Box 5] Results indicate substantial agreement that performance score can be used to distinguish quality → [Moderate]

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

Committee pre-evaluation comments

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

2a1. & 2b1. Specifications

Comments: **The data sources are administrative claims and health plan enrollment data. Specifications are clearly defined.
**Calculation is clear and data elements are clearly identified, codes are appropriate as is the logic evidence given is fairly unanimous that the higher the # of providers or pharmacies used the higher the risk to patients

2a2. Reliability Testing

Comments: ** The reliability testing level is the measure score. Several data sets were used -- Part D plan sample, one commercial health plan, and 8 state-based PDPs for Medicaid. Signal-to-noise analysis was conducted using a beta-binomial model. The mean reliability score was 0.9355 -- signaling adequate reliability.
**Signal to noise analysis of measure was completed with mean score of. 9355

2b2. Validity Testing

Comments: ** Validity testing was done at the measure score level; face validity only was done. The developer used its own approach to measure development and testing. Out of 12 member of the PQA's quality metrics expert panel, 67% agreed the measure reflects quality of care.
**The process of conducting face validity is not objective since it was done by the same group who developed the measure.
**Only face validity was performed - not unanimous agreement and also it appears that none of the content experts are actual healthcare providers who would be prescribing the drugs which is a weakness

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments:
**The exclusions are supported.
**There is no risk adjustment.
**Given the range of performance in each type of plan evaluated, the measure identifies meaningful differences about quality.
**No concerns about missing data.
**Other chronic conditions besides cancer might need to be excluded -- HIV or Sickle cell

Criterion 3. Feasibility

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.

- This measure is generated or collected by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims) Other data elements include Prescription claims data.
- ALL data elements are in defined fields in electronic claims.
- Pilot test sites indicated the measure was feasible and results were able to be reported efficiently and accurately. CMS calculates the measure for Part D plans. The prescription claims and medical data is readily available.
- Certain uses of the Measures are only approved with a licensing agreement from the developer, that specifies the terms of use and the licensing fee. The developer reserves the right to determine the conditions under which it will approve and/or license the Measures.

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

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **Feasibility is high since the data for this measure is generated by someone other than the person obtaining original information and is available from electronic claims. **No concerns with feasibility other than providers who don't use EHR how will their data be captured.

Criterion 4: Usability and Use

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

Current uses of the measure:

Publicly reported? ☐ **Yes** ☒ **No**

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

OR

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

Accountability program details:

- The measure was developed in 2015.
- The measure is currently being used in the Medicare Part D Overutilization Monitoring System to monitor the utilization of opioids for members with the Medicare drug benefit.
- CMS has announced plans to move this measure into the 2019 Part D Display Measures, using data from 2017.
- Reporting of results is not yet available.

Improvement results:

- There are no improvement results, as this is the initial endorsement submission.

Unexpected findings (positive or negative) during implementation:

- Developer did not identify any specific unexpected findings related to this measure.

Potential harms:

- Although no unintended negative consequences to individuals or populations were identified during testing, ,

concerns have been raised that prescribing changes such as dose reduction (without offering or arranging evidence-based treatment for patients with opioid use disorder) might be associated with unintended negative consequences, such as patients seeking heroin or other illicitly obtained opioids (1,2) or interference with appropriate pain treatment.(3) Data indicate that if access to prescription opioids is limited, some users of opioid analgesics will transition to heroin or other illicitly obtained opioids, leading to increased overdose death coincident with prescribing restrictions.(1) There are also concerns about pain being underdiagnosed and undertreated, particularly for ethnic and racial minorities, females, children or infirm elderly, or individuals who may be perceived to have mental health problems or are drug seeking.(4,5)

- These concerns must be balanced by the current situation in the United States which has been described by the CDC as an epidemic of opioid abuse, overdose, and deaths. Overdose involving opioid analgesics killed almost 17,000 persons in 2010 and the number of people with opioid analgesic use disorders increased to nearly 2 million.(6) The developer believes the potential benefits of monitoring those patients receiving the very highest doses of opioids for extended periods of time or receiving these drugs from multiple providers outweighs potential negative consequences.

Feedback :

- Developer did not identify any specific feedback loops related to this 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

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **This measure is not currently publicly reported or used in an accountability program. CMS has announced plans to move this measure into the 2019 Part D display measures.
**Measure is already being used in Medicare Part D overutilization monitoring although not publicly reported -- planned for 2019 display measure with 2017 data . This is initial endorsement and no other reporting or improvement results available

Criterion 5: Related and Competing Measures

Related or competing measures

- The measure is related to 2940 and 2951 which are being proposed for endorsement.

Harmonization

- N/A

Pre-meeting public and member comments

-

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Use of Opioids from Multiple Providers in Persons Without Cancer
IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: 5/13/2016

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus
The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.
4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.
5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.
6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)
Outcome
☐ Health outcome: Click here to name the health outcome

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

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

☒ Process: Prescriptions for opioids from multiple prescribers and multiple pharmacies

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

It has been shown that the measured process, prescriptions for opioids from multiple prescribers and pharmacies, correlates with undesired health outcomes. Use of multiple prescribers and pharmacies are associated with increased risks for opioid overdose. The risk for overdose increases with the number of prescribers and pharmacies.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☒ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

_____

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

_____

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

_____

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: January 2008 through August 2014

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for each new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

_____

1a.8 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.8.1 What process was used to identify the evidence?

A PubMed search was conducted using combinations of the following search terms: opioid, overdose, doctor shopping, pharmacy shopping, multiple prescribers, multiple pharmacies. Articles referenced in the identified articles were scanned for relevance. The CDC Guideline and Clinical and Contextual Evidence Reviews were also reviewed for relevant references (CDC Guideline: Dowell D, Haegerich TM, Chou R. CDC Guideline for Prescribing Opioids for Chronic Pain - United States, 2016. MMWR Recomm Rep. 2016 Mar 18;65(1):1-49. doi: 10.15585/mmwr.rr6501e1. Available at: http://www.cdc.gov/drugoverdose/prescribing/guideline.html.; CDC Clinical Evidence Review. Available at: http://stacks.cdc.gov/view/cdc/38026; CDC Contextual Evidence Review. Available at: http://stacks.cdc.gov/view/cdc/38027).

1a.8.2. Provide the citation and summary for each piece of evidence.

1. Dowell D, Haegerich TM, Chou R. CDC Guideline for Prescribing Opioids for Chronic Pain - United States, 2016. *MMWR Recomm Rep.* 2016;65(1):1-49. doi: 10.15585/mmwr.rr6501e1. Available at: http://www.cdc.gov/drugoverdose/prescribing/guideline.html.
   Summary: Recommendation 9 of the CDC Guidelines states that clinicians should review the patient's history of controlled substance prescriptions using state prescription drug monitoring program (PDMP) data to determine whether the patient is receiving opioid dosages or dangerous combinations that put him or her at high risk for overdose. Clinicians should review PDMP data when starting opioid therapy for chronic pain and periodically during opioid therapy for chronic pain, ranging from every prescription to every 3 months (recommendation category: A, evidence type: 4). The text related to recommendation 9 in the CDC guidelines states that although evidence is limited on the effectiveness of PDMP implementation at the state level on prescribing and mortality outcomes, most fatal overdoses have been shown to be associated with patients receiving opioids from multiple prescribers and/or with patients receiving high total daily opioid dosages; and information on both of these risk factors for overdose are available to prescribers in the prescription drug monitoring program (PDMP)." "Clinicians should discuss safety

concerns, including increased risk for respiratory depression and overdose, with patients found to be receiving opioids from more than one prescriber or receiving medications that increase risk when combined with opioids (e.g., benzodiazepines) and consider offering naloxone (see Recommendation 8)."

2. Yang Z, Wilsey B, Bohm M, Weyrich M, Roy K, Ritley D, Jones C, Melnikow J. Defining risk of prescription opioid overdose: pharmacy shopping and overlapping prescriptions among long-term opioid users in medicaid. *J Pain.* 2015;16(5):445-53. doi: 10.1016/j.jpain.2015.01.475. Epub 2015 Feb 11. PubMed PMID: 25681095.
Summary: An analysis of multistate Medicaid claims database (2008-2010) was conducted to evaluate strategies for identifying patients at high risk for overdose among enrollees who used 3 or more opioid prescriptions for 90 or more days. Diagnostic odds ratios were compared for opioid overdose events of 9 pharmacy shopping definitions. The diagnostic odds ratio for the criterion of 4 or more pharmacies in a 90-day period had the highest value at 5.40. The percentage of patients with opioid overdose events increased as the number of pharmacies increased.

3. Gwira Baumblatt JA, Wiedeman C, Dunn JR, Schaffner W, Paulozzi LJ, Jones TF. High-risk use by patients prescribed opioids for pain and its role in overdose deaths. *JAMA Intern Med* 2014;174(5):796–801. PMID: 24589873.
Summary: A matched case-control study among patients prescribed opioids in Tennessee (2008-2011) found an increased risk of opioid-related overdose death with 4 or more prescribers (adjusted odds ratio [aOR], 6.5; 95% Confidence Interval [CI], 5.1-8.5), 4 or more pharmacies (aOR, 6.0; 95% CI, 4.4-8.3), and more than 100 morphine milligram equivalents (MMEs) per day (aOR, 11.2; 95% CI, 8.3-15.1) daily mean dose. At least one of these risk factors was present in 55% of overdose deaths. Risk of overdose death increased with increasing number of pharmacies ($P < .001$) and prescribers used by the patient ($P < .001$).

4. Cepeda MS, Fife D, Chow W, Mastrogiovanni G, Henderson SC. Assessing opioid shopping behaviour: a large cohort study from a medication dispensing database in the US. *Drug Saf.* 2012;35(4):325-34. doi: 10.2165/11596600-000000000-00000. PMID: 22339505.
Summary: A cohort study of prescription data was conducted to provide a definition of shopping behavior that differentiates opioids from benzodiazepines and diuretics, avoiding the inappropriate flagging of individuals with legitimate use of opioids. The authors concluded that having 2 or more overlapping prescriptions written by different prescribers and filled at 3 or more pharmacies differentiates opioids from diuretics and likely constitutes shopping behavior.

5. Paulozzi LJ, Kilbourne EM, Shah NG, Nolte KB, Desai HA, Landen MG, Harvey W,Loring LD. A history of being prescribed controlled substances and risk of drug overdose death. *Pain Med.* 2012;13(1):87-95. doi: 10.1111/j.1526-4637.2011.01260.x. PMID: 22026451.
Summary: A matched case-control study in New Mexico (2006-2008) showed that risk of unintentional overdose death increased with the number of prescribers and pharmacies. The odds ratio of one more prescriber and pharmacy in the previous six months was 1.7 (95% CI, 1.6-1.9) and 2.3 (95% CI, 2.0-2.5), respectively.

6. Peirce GL, Smith MJ, Abate MA, Halverson J. Doctor and pharmacy shopping for controlled substances. *Med Care.* 2012;50(6):494-500. doi: 10.1097/MLR.0b013e31824ebd81. PubMed PMID: 22410408.
Summary: In a case-control study in West Virginia (2005-2007), subjects classified as doctor shoppers (4 or more prescribers in the previous 6 months) had 2 times the odds (OR, 2.0; 95% CI, 1.6–2.6), and pharmacy shoppers (4 or more pharmacies in the previous 6 months) had 3 times the odds (3.2; 95% CI, 2.3–4.5) of drug-related death compared with those classified as non-shoppers. Subjects classified as both doctor and pharmacy shoppers also had increased odds (OR, 3.6; 95% CI, 2.7–4.7) of drug-related death compared with non-shoppers.

7. Gomes T, Mamdani MM, Dhalla IA, Paterson JM, Juurlink DN. Opioid dose and drug-related mortality in patients with nonmalignant pain. *Arch Intern Med.* 2011;171(7):686-91. doi: 10.1001/archinternmed.2011.117. PubMed PMID: 21482846.
Summary: In an observational nested case-control study of decedents in Ontario Canada (1997-2006) who received an opioid prescription, those whose deaths were related to opioids were more likely to have obtained opioids from multiple physicians and pharmacies than decedents whose deaths were not related to opioids.

8. Katz N, Panas L, Kim M, Audet AD, Bilansky A, Eadie J, Kreiner P, Paillard FC, Thomas C, Carrow G. Usefulness of prescription monitoring programs for surveillance--analysis of Schedule II opioid prescription data in Massachusetts, 1996-2006. *Pharmacoepidemiol Drug Saf.* 2010;19(2):115-23. doi:10.1002/pds.1878. PMID: 20014166.
Summary: An analysis of Massachusetts prescription monitoring program (PMP) data was conducted to evaluate trends in opioid prescribing, dispensing, and usage of prescription data. The authors selected the criterion of 4 or more prescribers and 4 or more pharmacies as an indicator of potential non-medical use and diversion of prescription opioids such as doctor shopping. The authors commented that the criterion of 3 or prescribers and 3 or more pharmacies is not stringent enough and likely to misclassify patients who are using opioids appropriately (false positives).

9. White AG, Birnbaum HG, Schiller M, Tang J, Katz NP. Analytic models to identify patients at risk for prescription opioid abuse. *Am J Manag Care.* 2009;15(12):897-906. PubMed PMID: 20001171. Available at: http://www.ajmc.com/journals/issue/2009/2009-12-vol15-n12/AJMC_09Dec_White_897to906/
Summary: A retrospective observational study of privately insured patients in Maine (2005-2006) showed an increased risk for opioid dependence, abuse, or overdose among persons receiving multiple prescriptions, having multiple prescribers, or using multiple pharmacies.

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
2_PQA-Opioids_Multi_Provider_Evidence_Form_051016.docx

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., the benefits or improvements in quality envisioned by use of this measure)*
Abuse and overdose of prescription drugs is a major public health issue in the United States.(1,2) Studies have shown that people who see multiple prescribers or use multiple pharmacies are more likely to die of drug overdoses.(3)

PQA developed 3 measures related to prescription opioid use that are indicative of the quality of care for patients taking these medications. The measures examine the quality of use related to the dose of the medications over time, access to the medications through multiple providers, and the combination of both these criteria. This measure, Use of Opioids from Multiple Providers in Persons Without Cancer, focuses specifically on the use of opioids from multiple providers.

Prescription drug monitoring programs (PDMP), which track the use of multiple providers by patients, indicate that such use is typically found among a small proportion of patients, with the proportion declining as the number of providers increases. In Massachusetts in 2006, considering only Schedule II opioids, 0.5% of patients saw 4+ prescribers and 4+ pharmacies.(4) A national study found that 13% of patients had overlapping prescriptions from two or more different prescribers during an 18-month period. Of these, 0.5% used 4+ prescribers and 4+ pharmacies.(5) People who see multiple prescribers or use multiple pharmacies are more likely to die of drug overdoses.(3) When comparing the diagnostic odds ratio for opioid overdose events of 9 pharmacy shopping definitions, a threshold of 4 pharmacies had the highest diagnostic odds ratio.(6) Data from the California PDMP indicates that people with higher daily dosages are more likely to see multiple prescribers or go to multiple pharmacies.(7) However, there is no clear threshold at which multiple prescribers and multiple pharmacies represent lack of continuity or poorly coordinated care.

Data suggest that efforts to prevent opioid overdose deaths should include a multi-faceted approach focused on strategies that target high-dose opioid users as well as persons who seek care from multiple doctors and pharmacies. The data also suggests that these criteria can be considered separately, as measures related to prescribed opioids for appropriate clinical uses versus

inappropriate uses. Thus, as stated above, PQA developed 3 measures: one for high dose therapy, one for multiple providers, and one that is the intersection of both high dose and multiple providers – with this measure presently under consideration focused specifically on the use of opioids from multiple providers.

References:
1. U.S. Department of Health and Human Services, Office of Disease Prevention and Health Promotion. (2014). National Action Plan for Adverse Drug Event Prevention. Washington, DC. Accessed on: 4/9/15. Available at: http://www.health.gov/hcq/pdfs/ADE-Action-Plan-508c.pdf.
2. Liu Y, Logan JE, Paulozzi LJ, et al. Potential misuse and inappropriate prescription practices involving opioid analgesics. Am J Manag Care. 2013;19:648-65. PMID: 24304213.
3. Paulozzi LJ, Kilbourne EM, Shah NG, et al. A history of being prescribed controlled substances and risk of drug overdose death. Pain Med. 2012;13:87-95. PMID: 22026451.
4. Katz N, Panas L, Kim M, et al. Usefulness of prescription monitoring programs for surveillance--analysis of Schedule II opioid prescription data in Massachusetts, 1996- 2006. Pharmacoepidemiol Drug Saf. 2010;19:115-23. PMID: 20014166.
5. Cepeda MS, Fife D, Chow W, et al. Assessing opioid shopping behaviour: a large cohort study from a medication dispensing database in the US. Drug Saf. 2012;35:325-34. PMID: 22339505.
6. Yang Z, Wilsey B, Bohm Michele, et. al. Defining Risk of Prescription Opioid Overdose: Pharmacy Shopping and Overlapping Prescriptions Among Long-term Opioid Users in Medicaid. The Journal of Pain. 2015;445—453. PMID 25681095.
7. Han H, Kass PH, Wilsey BL, et al. Individual and county-level factors associated with use of multiple prescribers and multiple pharmacies to obtain opioid prescriptions in California. PLoS One. 2012;7:e46246. PMID: 23049992.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
The measure was tested in three different health plan data sources – the Medicare population, one commercial heath plan, and the Medicaid population.

The testing from the Medicare population used administrative claims data from January 1st 2013 to December 31st 2013. For the Medicare testing, the analysis included a convenience sample of over 700 Medicare Part D prescription drug plans, covering 7,067,445 individuals aged 18 and older.

The Medicare rates ranged from 17.98 per 1,000 to 31.00 per 1,000. The mean was 23.31 per 1,000 and the median was 26.12 per 1,000. The standard deviation was 5.73. The 25th percentile was 22.32 per 1,000, the 50th percentile is the median (26.12 per 1,000) and the 75th percentile was 29.10 per 1,000. The interquartile range was 6.78.

The majority of testing used Medicaid prescription claims data from January 1st 2015-December 31st 2015. Testing also included prescription claims data from one state's Medicaid plan from July 1st 2014-June 30th 2015. Testing included 8 state based prescription drug plans in 6 states, covering 1,437,410 individuals age 18 and older.

The Medicaid rates ranged from 57.56 per 1,000 to 93.08 per 1,000. The mean was 72.28 per 1,000 and the median was 69.93 per 1,000. The standard deviation was 12.03. The 25th percentile was 65.13 per 1,000, the 50th percentile is the median (69.93 per 1,000) and the 75th percentile was 80.66 per 1,000. The interquartile range was 15.51.

Testing was also conducted in one Commercial health plan using administrative claims from January 1st 2013 to December 31st 2013. This plan covered 209,191 individuals age 18 and older. The measure rate for this plan was 20.57 per 1,000.

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*
Disparities data is available for the Medicare population. The testing from the Medicare population used administrative claims data

from January 1st 2013 to December 31st 2013 and included a convenience sample of over 700 Medicare Part D prescription drug plans, covering 7,067,445 individuals aged 18 and older.

The beneficiary level Low Income Subsidy (LIS) variable was used to determine disparities in rates for populations with different sociodemographic status. The LIS is a subsidy paid by the Federal government to the drug plan for Medicare beneficiaries who need extra help with their prescription drug costs due to limited income and resources. The measure rate for the LIS group is 42.42 per 1,000 while the rate for the non-LIS population is significantly lower, at 11.71 per 1,000.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
N/A

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, A leading cause of morbidity/mortality, Patient/societal consequences of poor quality
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
The misuse of prescription opioids in America is a public health crisis and addressing the overdose epidemic is a high priority for the US government.(1-4) Deaths from drug overdose have risen steadily over the past two decades and have become the leading cause of injury death in the United States.(5) Since 1999, prescription opioid use and overdose deaths have quadrupled.(6) More than 165,000 people have died from prescription opioids in this timeframe,(7) yet there has not been an overall change in the amount of pain that Americans report.(8,9) In 2014, more than 14,000 people died from prescription opioid overdose, more than any year on record.(7) Use of multiple prescribers and pharmacies are associated with increased risks for opioid overdose.(10-13) The risk for overdose increases with the number of prescribers and pharmacies. Identifying patients at higher risk for overdose is a component of improving opioid prescribing practices to improve patient safety.(2)

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
1. S.524 - Comprehensive Addiction and Recovery Act of 2016. Available at: https://www.congress.gov/bill/114th-congress/senate-bill/524/text.
2. Dowell D, Haegerich TM, Chou R. CDC Guideline for Prescribing Opioids for Chronic Pain - United States, 2016. MMWR Recomm Rep. 2016;65:1-49. doi:10.15585/mmwr.rr6501e1. (PMID: 26987082).
3. HHS. ASPE Issue Brief: Opioid Abuse in the U.S. and HHS Actions to Address Opioid-Drug Related Overdoses and Deaths; 2015. Available at: http://aspe.hhs.gov/basic-report/opioid-abuse-us-and-hhs-actions-address-opioid-drug-related-overdoses-and-deaths.
4. US Department of Health and Human Services. National Action Plan for Adverse Drug Event Prevention. Washington, DC; 2014. Available at: http://health.gov/hcq/ade.asp.
5. CDC. Web-based Injury Statistics Query and Reporting System (WISQARS). 2014. Available at: http://www.cdc.gov/injury/wisqars/fatal.html.
6. Chen LH, Hedegaard H, Warner M. Drug-poisoning Deaths Involving Opioid Analgesics: United States, 1999-2011. NCHS Data Brief. 2014;(166):1-8
7. CDC. Wide-ranging online data for epidemiologic research (WONDER). Atlanta, GA: CDC, National Center for Health Statistics; 2016. Available at: http://wonder.cdc.gov.
8. Chang H, Daubresse M, Kruszewski S, et al. Prevalence and treatment of pain in emergency departments in the United States, 2000 – 2010. Amer J of Emergency Med 2014; 32(5): 421-31.
9. Daubresse M, Chang H, Yu Y, Viswanathan S, et al. Ambulatory diagnosis and treatment of nonmalignant pain in the United States, 2000 – 2010. Medical Care 2013; 51(10): 870-878.
10. Gwira Baumblatt JA, Wiedeman C, Dunn JR, Schaffner W, Paulozzi LJ, Jones TF. High-risk use by patients prescribed opioids for pain and its role in overdose deaths. JAMA Intern Med 2014;174(5):796–801. (PMID: 24589873.) ?
11. Paulozzi LJ, Kilbourne EM, Shah NG, Nolte KB, Desai HA, Landen MG, Harvey W,Loring LD. A history of being prescribed controlled

substances and risk of drug overdose death. Pain Med. 2012;13(1):87-95. doi: 10.1111/j.1526-4637.2011.01260.x. PMID: 22026451.
12. Peirce GL, Smith MJ, Abate MA, Halverson J. Doctor and pharmacy shopping for controlled substances. Med Care. 2012;50(6):494-500. doi: 10.1097/MLR.0b013e31824ebd81. PubMed PMID: 22410408.
13. Gomes T, Mamdani MM, Dhalla IA, Paterson JM, Juurlink DN. Opioid dose and drug-related mortality in patients with nonmalignant pain. Arch Intern Med. 2011;171(7):686-91. doi: 10.1001/archinternmed.2011.117. PubMed PMID: 21482846.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
N/A

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:**
Measure Title Use of Opioids from Multiple Providers in Persons Without Cancer
Date of Submission: 5/13/2016
Type of Measure:

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

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and**

composite performance measures, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decision making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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.23*)**	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ administrative claims	☒ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

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

The measure was tested in three different health plan data sources – the Medicare population, one commercial heath plan, and the Medicaid population.

For the Medicare population, data used for testing came from three different sources. The Medicare Part D Prescription Drug Event (PDE) claims were used for the identification of prescription drugs and cancer exclusions. To identify dates of birth and continuous enrollment, the Common Medicare Environment (CME) data source was used. To identify hospice enrollment, the Medicare Enrollment Database (EDB) was used.

The data source for the Commercial population came from the health plans' enrollment data, medical claims, and prescription claims.

For the Medicaid population, the data used for testing came from Medicaid administrative claims. Six Medicaid plans covering four states were included in the testing using data from a Pharmacy Benefits Manager (PBM) organization. In addition, two other state-based plans were included in the testing using their state Medicaid administrative claims database. Medical claims were used to identify the cancer diagnoses, and the pharmacy claims were used for the identification of prescription drugs.

1.3. What are the dates of the data used in testing? Click here to enter date range

The testing from the Medicare and Commercial populations used administrative claims data from January 1st 2013 to December 31st 2013. The majority of testing used Medicaid prescription claims data from January 1st 2015-Decemer 31st 2015. The data from this time period were the most complete recent data available at the time of testing. Testing also included prescription claims data from one state's Medicaid plan from July 1st 2014-June 30th 2015.

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

Measure Specified to Measure Performance of: **(*must be consistent with levels entered in item S.26*)**	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☒ health plan	☒ health plan
☐ other: Click here to describe	☐ other: Click here to describe

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

For the Medicare testing, the analysis included a convenience sample of over 700 Medicare Part D prescription drug plans.

Testing was also conducted in one Commercial health plan. The size and characteristics of these populations are included at the patient level in 1.6.

For the Medicaid testing, the analysis included 8 state based prescription drug plans covering 6 states. 3 plans were from the same state in the Mid-Atlantic region of the United States (US), 2 plans were from states in the South Atlantic region of the US, two plans were from states in the West South Central region of the US, and one plan was from a state in the East South Central region of the US. The size and characteristics of the population are included at the patient level in 1.6.

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

For the Medicare testing, a total of 7,067,445 individuals aged 18 and older were included in the testing and analysis. This data can be stratified by age, gender, and type of Part D plan. Of all persons, 2,531,712 (35.8%) are male, and 4,535,732 (64.2%) are female. Individuals by age group included 271,635 (3.8%) age 18-40, 2,159,384 (30.6%) age 41-64 and 4,636,425 (65.6%) over age 65. Of all individuals, 2,492,658 (35.3%) are enrolled in a Medicare Advantage Prescription Drug Plan (MA-PD) and 4,574,787 (64.7%) are enrolled in a standalone Prescription Drug Plan (PDP).

For the Commercial plan, a total of 209,191 individuals age 18 and older were included in the analysis. Of all persons 92,227 (44.1%) are male, and 116,964 (55.9%) are female. Persons by age group included 46,913 (22.4%) age 18-40, 133,207 (63.7%) age 40-64 years, and 29,071 (13.9%) age 65 and older.

For the Medicaid plans, a total of 1,437,410 individuals age 18 and older were included in the analysis. Of all persons 515,164 (35.8%) are male, and 922,246 (64.2%) are female. Persons by age group included 897,641 (62.4%) age 18-40, 454,528 (31.6%) age 40-64 years, and 85,241 (6.0%) age 65 and older.

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

The Medicaid data was used to test reliability. This data does not include the RxHCC indicator to identify cancer exclusions, and instead uses ICD-9 or ICD-10 (depending on the year of the data) to identify diagnostic criteria for the cancer exclusions. The Medicaid data also does not allow for identification of hospice patients.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

For the Medicare population, the beneficiary level Low Income Subsidy (LIS) variable was used to determine disparities in rates for populations with different sociodemographic status. The LIS is a subsidy paid by the Federal government to the drug plan for Medicare beneficiaries who need extra help with their prescription drug costs due to limited income and resources. For the Commercial and Medicaid other populations, no patient level indicators of sociodemographic status were available in the data.

2a2. RELIABILITY TESTING

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

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

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

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

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

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

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

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

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

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

Using the parameter estimates from the Beta-Binomial model we computed individual plan reliability scores. Table 1 below shows the distribution of the plan-level scores. Plans have very high reliability scores. The reliability score mean is 0.9355 and the median 0.9518.

Table 1. Individual Plan Reliability Score Distribution

Statistic	Values
Mean	0.9355
Standard Dev.	0.0621
Min	0.7911
p10	0.8817
p25	0.9330
p50 (Median)	0.9518
p75	0.9728

p90	0.9769
max	0.9863

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

The reliability score mean is 0.9355 and the median 0.9518. A reliability score of 0.7 is the minimum threshold for reliability. Based on the high reliability scores for each of the plans in the analysis, the measure is considered reliable.

_____

2b2. VALIDITY TESTING

2b2.1. 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*)

2b2.2. 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)*

PQA uses a systematic, transparent, consensus-based measure development and testing process. That process used in 2014 to develop this measure is outlined below:

- Step 1: PQA workgroups identify measure concepts that may be appropriate for development into fully specified performance measures. The workgroups focus on specific aspects of the medication-use system and/or specific therapeutic areas. The workgroups are open to all members of PQA and use a consensus-based approach to identify, prioritize and recommend the measure concepts that are deemed to be highly important for supporting quality improvement related to medications.

- Step 2: The measure concepts that are recommended for further development through a vote by the PQA workgroups are forwarded to the PQA Quality Metrics Expert Panel (QMEP) for evaluation and refinement. The QMEP is composed of PQA members who have backgrounds in pharmacy, medicine, research, quality improvement and measures development. The QMEP reviews the measure concepts to provide an initial assessment of the key properties of performance measures (i.e., feasibility, usability and scientific validity). The measure concepts that are rated highly on these key properties will then undergo technical specification.

- Step 3: The draft measure is provided to PQA member organizations for their comments prior to preparing technical specifications for pilot testing. The QMEP reviews member comments, edits the draft measure accordingly and poses testing questions based on this all-member feedback.

- Step 4: PQA selects partners to test the draft measure. These partners are often PQA member health plans or

academic institutions with expertise in quality and performance measure testing. The testing partner implements the draft technical specifications with their existing datasets and provides a report to PQA that details testing results and recommendations for modifications of the technical specifications.

- Step 5: The workgroup that developed the measure reviews the testing results and provides comment. The QMEP reviews the workgroup comments, testing results, recommendations and potential modifications and provides a final assessment of the feasibility and scientific validity of the draft performance measures.

- Step 6: Measures that are recommended by the QMEP for endorsement are posted on the PQA web site for member review, written comments are requested, and a conference call for member organizations is scheduled to address any questions. This process allows members to discuss their views on the measures in advance of the voting period.

- Step 7: PQA member organizations, which include organizations such as large pharmacy chains, health plans, quality organizations and pharmaceutical companies vote on the performance measure(s) considered for approval and/or endorsement.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

The measure was assessed for face validity (i.e., whether it appears to measure what it intends to measure) through review by the PQA workgroup that developed the measure, the PQA Quality Metrics Expert Panel (QMEP), and PQA's full membership. In addition, feedback about validity of the measure was sought out by the five PQA member organizations who tested the measure using their own data.

The PQA Medication Use Safety Workgroup was composed of 72 PQA members that worked on multiple measure concepts. After the workgroup completed the development of the measure specifications, 37 members of the workgroup voted to determine if the draft measure should continue on further development and review by the PQA QMEP. 94.6% of members recommended that the measure move on for QMEP review.

The PQA QMEP is a panel that includes individuals with expertise and experience in pharmacy, medicine, research, and clinical or other technical expertise related to quality improvement and measure development. The names and credentials of the QMEP Panel are listed in Table 1. The QMEP reviewed the measure prior to testing to ensure scientific soundness and usefulness. The QMEP reviewed the results of the measure testing including the performance measure scores reported by plan referenced in Section 2b5 (below). Out of the 12 members of the QMEP who voted, 67% strongly agreed that the measure results reflected the quality of care, and recommended that the measure be considered for endorsement by the PQA membership.

Table 1. **PQA Quality Metrics Expert Panel (QMEP)**

QMEP Member Name and Credentials	QMEP Member Organization
Bimal Patel, Pharm D, MS	MedImpact
Catherine Coast, PharmD	Highmark
Chris DuPaul, MBA	CVS Caremark
Christopher Dezii, RN, MBA, CPHQ	Bristol-Myers Squibb
Christopher Powers, PharmD	CMS
David Nau, PhD, RPh, CPHQ	Pharmacy Quality Solutions

QMEP Member Name and Credentials	QMEP Member Organization
Gary Erwin, PharmD	OmniCare
Gary Young, JD, PhD	Northeastern University
Jenny Weber, PharmD, MS, PCPS,CGP, BCACP	Humana
Jessica Frank, PharmD	OutcomesMTM
Karen Farris, PhD	University of Michigan
Keith Widmer, RPh, BCPP	Express Scripts
Kent Summers, RPh, PhD	Astellas
Lynn Deguzman, PharmD, CGP	Kaiser Permanente
Mary Ann Kliethermes, PharmD	Midwestern University
Mitzi Wasik, PharmD, PCPS	Coventry Health Care/Aetna
Pat Gleason, Pharm D, BCPS	Prime Therapeutics
Steve Riddle, PharmD, BCPS	Wolters Kluwer Health
Steven Burch, RPh, PhD	GlaxoSmithKline
Tony Willoughby, PharmD	HealthMart-McKesson

PQA membership was notified prior to the PQA Annual Meeting in May 2015, of the opportunity to consider and vote for the performance measure during the meeting. (Note: PQA membership comprises health plans, community pharmacy, long-term care pharmacies, HIT companies, PBMs, healthcare quality and standards organizations, professional and trade associations, and others.) Members received the measure description, key points and evidence, measure specifications, and the performance measure scores reported by plan. During the PQA Business meeting, the measure was reviewed. Nearly all of PQA membership had a representative at the Annual Meeting and were present for the vote. Voting options included, “Agree” (indicating that the organization approved the measure), “Disagree (indicating that the organization opposed the measure) and “Abstain.” Out of the 90 number of PQA members who participated in voting, 78.9% of the membership voted in favor of endorsing the measure.

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

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

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

_____

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

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

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

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

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

2b3.2. 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*)

Of the eight health plans included in the analysis, the cancer patient exclusions were 0.5% to 1.9% of the overall population. The one Medicaid plan that could identify hospice exclusions found only 15 cases, which represented 0.003% of their total population.

2b3.3. 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 results show that in some plans, almost 2% of the population has cancer and would be included in the measure if cancer was not excluded. This is a significant proportion of the population that could potentially impact the measure rates. No inferences about the hospice exclusion could be drawn because the majority of the plans could not identify exclusions.

2b4. 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 2b5.

2b4.1. What method of controlling for differences in case mix is used?
☐ **No risk adjustment or stratification**
☐ **Statistical risk model with** Click here to enter number of factors **risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use 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.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

2b4.4a. What were the statistical results of the analyses used to select risk factors?

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

2b4.5. 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 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

2b4.10. 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*)

2b4.11. 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*)

_____

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

To assess significant differences in measure rates, the data described in sections 1.5 and 1.6 above were used to calculate the mean, median, standard deviation, and interquartile range for the measure rates for the Medicare population and the Medicaid population. In addition, for the Medicaid population, the rates were divided into quartiles, and a Student's t-test was used to compare the rates between the bottom quartile (75th percentile) and top quartile (25th percentile). A student's t-test was used to compare the rates of the plans in the 25th percentile to the plans with rates in the 75th percentile. The statistics are for the Medicare population is reported below in 2b5.2, Tables 1 and 2. The statistics for the Medicaid population is reported below in 2b5.2, Tables 3 and 4.

2b5.2. 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 1. Variation in Measure Rates - Medicare Population (reported as number per 1,000 members)

Mean	Median	Standard Deviation
23.31	26.12	5.73

Table 2. Interquartile Range of Measure Rates - Medicare Population (reported as number per 1,000 members)

Minimum	17.98
25th Percentile	22.32
50th Percentile	23.12
75th Percentile	29.10
Maximum	31.00
Interquartile Range	6.78

Table 3. Variation in Measure Rates - Medicaid Population (reported as number per 1,000 members)

Mean	Median	Standard Deviation
72.28	69.93	12.03

Table 4. Interquartile Range of Measure Rates - Medicaid Population (reported as number per 1,000 members)

Minimum	57.56
25th Percentile	65.13
50th Percentile	69.93
75th Percentile	80.66
Maximum	93.08
Interquartile Range	15.51
Student's t-test p-value	0.034

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do*

the results mean in terms of statistical and meaningful differences?)

For the Medicare population, the measure rates showed significant variation, with a standard deviation of 5.73 and an Interquartile Range of 6.78. For the Medicaid population, the measure rates showed greater variation, with a standard deviation of 12.03 and an Interquartile Range of 15.51. There is a statistically significant difference in measure rates between the top and bottom quartile of the plans included in the testing (*P*=0.034 at alpha=0.05). This variation shows that there are meaningful differences in rates across plans.

2b6. 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 SDS 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 SDS 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.***

Only one set of specifications is provided for this measure.

2b6.1. 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*)

2b6.2. 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*)

2b6.3. 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*)

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

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

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

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

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

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

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

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

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 subcriteria 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):*
Behavioral Health : Alcohol, Substance Use/Abuse, Mental Health : Alcohol, Substance Use/Abuse

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Overuse, Safety : Medication Safety

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.)*
http://pqaalliance.org/measures/default.asp

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

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

Attachment **Attachment:** Cancer_Exclusion_RxHCC-_ICD-9_and_10_Codes-635969250747751020.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
N/A

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Any member in the denominator who received opioid prescription claims from 4 or more prescribers AND 4 or more pharmacies.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
The measurement year

S.6. **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, 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.
For each member in the denominator:
1. Calculate the number of unique pharmacy providers associated with an opioid prescription claim.
2. Calculate the number of unique prescribers associated with an opioid prescription claim.
3. Any member with four or more unique pharmacy providers AND four or more unique prescribers meets the criteria for the Numerator.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Any member with two or more prescription claims for opioids filled on at least two separate days, for which the sum of the days supply is greater than or equal to 15.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, Populations at Risk : Dual eligible beneficiaries

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
Any member with two or more prescription claims for opioids filled on at least two separate days, for which the sum of the days supply is greater than or equal to 15.

Table Opioid-A: Opioid Medications

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

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Any member with a diagnosis for Cancer or a Prescription Drug Hierarchical Condition Category (RxHCC) 8, 9, 10, or 11 for Payment Year 2015; or RxHCC 15, 16, 17, 18, or 19 for Payment Year 2016; (see list in S.11 and S.2b); or a hospice indicator from the enrollment database.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Hospice Exclusion: Exclude those members identified in the Medicare Enrollment Database as being enrolled in hospice.

Cancer Exclusion: For Payment Year 2015: RxHCC 8, 9, 10, or 11. For Payment Year 2016: RxHCC 15, 16, 17, 18, or 19
ICD 9 and 10 Codes to Identify Cancer: Please see attachment in S2.b

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
The measure is stratified by the following lines of business for the health plan:

- Commercial
- Medicare
- Medicaid

Medicare Plans are further stratified by Low Income Subsidy status
Definition: Medicare Low Income Subsidy (LIS)
A subsidy paid by the Federal government to the drug plan for Medicare beneficiaries who need extra help with their prescription drug costs due to limited income and resources. Medicare beneficiaries apply for the LIS with the Social Security Administration or their State Medicaid agency.

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

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
N/A

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
N/A

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **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.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
Step One:
Calculate the denominator by identifying the number of all eligible members with two or more prescription claims for opioids filled on at least two separate days, for which the sum of the days supply is greater than or equal to 15.

Step Two:
Calculate the numerator by:
a. Calculate the number of unique pharmacy providers associated with an opioid prescription claim.

b. Calculate the number of unique prescribers associated with an opioid prescription claim.
c. Any member with four or more unique pharmacy providers AND four or more unique prescribers meets the criteria for the Numerator.

Step Three:
Divide the number of members that met the criteria in numerator (Step Two c.) by the denominator (Step One) and multiply times 1000. The rate is reported as a proportion: XX out of 1,000 members.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
N/A

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
N/A

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Health Plan Medical and Pharmacy Claims. Health Plan member enrollment information.

S.25. **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.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Health Plan, Population : National, Population : State

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Other, Pharmacy
If other: The level of analysis for this measure is the prescription drug health plan, but it contains claims data from multiple care settings, including ambulatory, skilled nursing facility, pharmacy etc.

S.28. **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

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
PQA_Multiprovider_testing_attachment.docx

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue

burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

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

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

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

If other: Prescription claims data

3b. Electronic Sources

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

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

ALL data elements are in defined fields in electronic claims

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

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.**

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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

Pilot test sites indicated the measure was feasible and results were able to be reported efficiently and accurately. CMS calculates the measure for Part D plans. The data is readily available (prescription claims data and medical data).

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

PQA develops and maintains numerous performance measures related to the medication use system. The Measures are the proprietary property of PQA, and it is in the interest of PQA to protect and promote the appropriate use of the Measures. PQA may approve an organization ´s use of the Measures; however, no organization may use the Measures without first obtaining permission from PQA prior to using the Measures. Certain uses of the Measures are only approved with a licensing agreement from PQA that specifies the terms of use an the licensing fee. PQA reserves the right to determine the conditions under which it will approve and/or license the Measures.

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. **Current and Planned Use**

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

Planned	Current Use (for current use provide URL)
Public Reporting Quality Improvement (Internal to the specific organization)	Quality Improvement with Benchmarking (external benchmarking to multiple organizations) CMS Medicare Part D - Patient Safety Reports http://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovGenIn/index.html

4a.1. **For each CURRENT use, checked above, provide:**

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

Name of program and sponsor: CMS Medicare Part D Drug Benefit
Purpose: Monitor Opioid use by Part D beneficiaries
Geographic area: National, approximately 38 million beneficiaries in Part D plans.

4a.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 was developed in 2015.

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
The measure is currently being used in the Medicare Part D Overutilization Monitoring System to monitor the utilization of opioids for members with the Medicare drug benefit.

CMS has announced plans to move this measure into the 2019 Part D Display Measures, using data from 2017.

Reporting of results is not yet available.

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

N/A

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
N/A

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
No unintended negative consequences to individuals or populations were identified during testing. This measure, Use of Opioids from Multiple Providers in Persons Without Cancer, has been implemented by CMS Part D as part of the Overutilization Monitoring System beginning January, 2016. To date, no negative consequences have been identified.

However, concerns have been raised that prescribing changes such as dose reduction (without offering or arranging evidence-based treatment for patients with opioid use disorder) might be associated with unintended negative consequences, such as patients seeking heroin or other illicitly obtained opioids (1,2) or interference with appropriate pain treatment.(3) Data indicate that if access to prescription opioids is limited, some users of opioid analgesics will transition to heroin or other illicitly obtained opioids, leading to increased overdose death coincident with prescribing restrictions.(1) There are also concerns about pain being underdiagnosed and undertreated, particularly for ethnic and racial minorities, females, children or infirm elderly, or individuals who may be perceived to have mental health problems or are drug seeking.(4,5)

These concerns must be balanced by the current situation in the United States which has been described by the CDC as an epidemic of opioid abuse, overdose, and deaths. Overdose involving opioid analgesics killed almost 17,000 persons in 2010 and the number of people with opioid analgesic use disorders increased to nearly 2 million.(6) We believe the potential benefits of monitoring those patients receiving the very highest doses of opioids for extended periods of time or receiving these drugs from multiple providers outweighs potential negative consequences.

References:
1. Coffin P, Banta-Green C. The dueling obligations of opioid stewardship. Ann Intern Med. 2014;160:207–8. doi.org/10.7326/ M13-2781. (PMID 25133372).
2. Cicero,T, Ellis M, Harney J. Shifting Patterns of Prescription Opioid and Heroin Abuse in the United States. N Engl J Med. 2015; 373:1789-90. DOI: 10.1056/NEJMc1505541. (PMID 26510045).
3. Twillman RK, Kirch R, Gilson A. Efforts to control prescription drug abuse: Why clinicians should be concerned and take action as essential advocates for rational policy. CA Cancer J Clin. 2014;64:369–76. doi.org/10.3322/caac.21243. (PMID 25044063).
4. Kirschner N, Ginsburg J, Snyder LS, Health and Public Policy Committee of the American College of Physicians. Prescription Drug Abuse; executive summary of a policy position paper from the American College of Physicians. Ann Intern Med. 2014;160:198-200. doi:10.7326/M13-2209. (PMID 24323199).
5. Institute of Medicine (US) Committee on Advancing Pain Research, Care, and Education. Relieving Pain in America: A Blueprint for Transforming Prevention,Care, Education, and Research. Washington (DC): National Academies Press (US); 2011. (PMID: 22553896). Available at: http://www.nap.edu/read/13172/chapter/1
6. Coffin P, Banta-Green C. The dueling obligations of opioid stewardship. Ann Intern Med. 2014;160:207. doi: 10.7326/M13-2781. (PMID 24322334).

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
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 completely harmonized?

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.)
N/A

Appendix

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

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** PQA
Co.2 **Point of Contact:** Julie, Kuhle, jkuhle@pqaalliance.org, 515-554-6685-
Co.3 **Measure Developer if different from Measure Steward:** PQA
Co.4 **Point of Contact:** Julie, Kuhle, jkuhle@pqaalliance.org, 515-554-6685-

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

PQA Quality Metrics Expert Panel 2015
Responsible for review and consideration of the measure concept and all testing results of the draft measure
Bimal Patel*, Pharm D, MS MedImpact
Catherine Coast, PharmD Highmark
Chris DuPaul, MBA CVS Caremark
Christopher Dezii,RN, MBA, CPHQ Bristol-Myers Squibb
Christopher Powers, PharmD CMS
David Nau, PhD, RPh, CPHQ PQS
Gary Erwin, PharmD Omnicare
Gary Young, JD, PhD Northeastern University
Jenny Weber*, PharmD, MS, PCPS,CGP, BCACP Humana
Jessica Frank, PharmD OutcomesMTM
Karen Farris, PhD University of Michigan

Keith Widmer, RPh, BCPP	Express Scripts
Kent Summers, RPh, PhD	Astellas
Lynn Deguzman, PharmD, CGP	Kaiser Permanente
Mary Ann Kliethermes, PharmD	Midwestern University
Mitzi Wasik, PharmD, PCPS	Coventry Health Care/Aetna
Pat Gleason, Pharm D, BCPS	Prime Therapeutics
Steve Riddle, PharmD, BCPS	Wolters Kluwer Health
Steven Burch, RPh, PhD	GlaxoSmithKline
Tony Willoughby, PharmD	HealthMart-McKesson

* denotes co-chair

PQA Medication Use Safety Workgroup 2014

Responsible for development of the measure

Amber Baybayan OutcomesMTM
David Belew MedHere Today
Rachael Boggs PQA Invited Guest Participant
Stay Bontha PerformRx
Sara Burnheimer UPMC Health Plan
Patrick Campbell University of Arizona College of Pharmacy
Scott Campbell PQA Invited Guest Participant
Rebecca Chater Ateb
Trina Clark GlaxoSmithKline
Victor Cohen American Society of Health-System Pharmacists (ASHP)
Michael Contos Indian Health Services
Karen Davidson Therapeutic Research Center (home of Pharmacist's Letter and Prescriber's Letter)
Shelly Delaville American Society of Consultant Pharmacists (ASCP)
James DeVita CVS/Caremark
Sara Ericsson MedImpact Healthcare Systems, Inc.
Marybeth Farquhar URAC
Alison Farrell Ahold USA
Cindi Fitzpatrick U.S. Food & Drug Administration (FDA)
Jeremy Fredell Express Scripts, Inc.
George Garmer CARE Pharmacies Cooperative
Jennifer Gatsos-Walter Wolters Kluwer Health, Clinical Solutions
Mary Ghods U.S. Food & Drug Administration (FDA)
James Glass Rite Aid
Averill Gordon Walgreen Co.
Lindsey Gumbo Pharmaceutical Research & Manufacturers of America (PhRMA)
Tracy Harrell SinfoníaRx
Tiffany Harris SCAN Health Plan
Shannon Harrison Highmark Health Services
Lisa Hines* University of Arizona College of Pharmacy
John Kessler National Alliance of State Pharmacy Associations (NASPA)
Mi´a Kirkland Wellcare
Nicholas Kostek Kaiser Permanente
Maribeth Kowalski Purdue Pharma, L.P.
Jason Kinsman RxAnte
Edward Lennard U.S. Office of Personnel Management
Patricia Marchlowska Lilly USA
Peter Marshall HealthPartners
Kevin Masci Target
Richard McLeod Pfizer, Inc.
Diane McNally Centers for Medicare & Medicaid Services (CMS)
Brent Merrick Cigna-HealthSpring
Leslie Miller Gorman Health Group
Joel Montavon Catamaran
Kim Moon PQA Invited Guest Participant

Gina Moore PQA Invited Guest Participant
Scott Nakagawa Applied Research Works
Patricia Neafsey ActualMeds Corporation
Jeffrey Nesheim Takeda Pharmaceuticals America, Inc.
Michael Nguyen CenseoHealth
Kyle Null* University of Mississippi Center for Pharmaceutical Marketing & Management
Udo Nwachukwu Mirixa Corporation
Steven Oh Health Mart Systems Inc.
Maria Osborne American Pharmacists Association (APhA)
Nicole Paterson Fairview Medication Therapy Management
Jacqui Pesa Johnson & Johnson
Roger Pinsonneault RelayHealth
Richard Segal University of Florida College of Pharmacy
Bupendra Shah Long Island University Arnold & Marie Schwartz College of Pharmacy
Christine Sommer First DataBank
Catherine Starner Prime Therapeutics
Karen Stockl UnitedHealth Group
Brian Sweet AstraZeneca
Christie Teigland Inovalon, Inc.
Jennifer Thomas National Alliance of State Pharmacy Associations (NASPA)
Ly Tran PharmMD
Maria Vassilakis Astellas Scientific and Medical Affairs, Inc.
Kathleen Vest PQA Invited Guest Participant
Brandi Rosberg Walmart
Jennifer Weber Humana
Elizabeth Whaley-Buono MeadWestvaco
Jennifer Williams Aetna
Melissa Wilson Capital Health Plan

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2015

Ad.3 **Month and Year of most recent revision:** 05, 2015

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

Ad.5 **When is the next scheduled review/update for this measure?** 10, 2016

Ad.6 **Copyright statement:** Rights Retained by PQA, Inc 2016.

Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 2951
De.2. **Measure Title:** Use of Opioids from Multiple Providers and at High Dosage in Persons Without Cancer
Co.1.1. **Measure Steward:** PQA
De.3. **Brief Description of Measure:** The proportion (XX out of 1,000) of individuals without cancer receiving prescriptions for opioids with a daily dosage greater than 120mg morphine equivalent dose (MED) for 90 consecutive days or longer, AND who received opioid prescriptions from four (4) or more prescribers AND four (4) or more pharmacies.
1b.1. **Developer Rationale:** Abuse and overdose of prescription drugs is a major public health issue in the United States.(1,2) Though there is no FDA maximum dose or duration for opioid drugs, studies have demonstrated that patient populations taking high opioid doses for prolonged periods are often characterized by high rates of psychiatric and substance abuse disorders, frequently do not receive care consistent with clinical guidelines, and have higher death rates.(3-6) Studies have shown that people who see multiple prescribers or use multiple pharmacies are more likely to die of drug overdoses.(6)

PQA developed 3 measures related to prescription opioid use that are indicative of the quality of care for patients taking these medications. The measures examine the quality of use related to the dose of the medications over time, and access to the medications through multiple providers. This measure, Use of Opioids from Multiple Providers in Persons Without Cancer, focuses on the use of opioids at high dose and from multiple providers.

Claims data from commercially insured patients indicate that approximately 8% of opioid prescriptions for acute pain and 12% for chronic pain specify a daily dosage of 120mg MED or more.(2) The Washington State Agency Medical Directors Group has suggested 120mg MED as a dosage level that should not be exceeded without special consideration.(4) Group Health Cooperative (GHC), which implemented this guidance from the 2010 edition, has demonstrated a reduction in opioid doses for their patients with chronic pain. For the last quarter of 2014, less than one-quarter of these patients seen by GHC providers received 50 mg/day MED or greater and only 7.3% exceeded 120 mg/day MED.(4) The proportion of patients being treated at this dosage for more than 90 days has not been described. However, one study of veterans treated with 180mg MED/day or more for 90+ days (3) found that this group was characterized by high rates of psychiatric and substance abuse disorders and frequently did not receive care consistent with clinical guidelines. Studies suggest that high opioid dosage increases the risk of overdoses and fractures.(5-7)

Prescription drug monitoring programs (PDMP), which track the use of multiple providers by patients, indicate that such use is typically found among a small proportion of patients, with the proportion declining as the number of providers increases. In Massachusetts in 2006, considering only Schedule II opioids, 0.5% of patients saw 4+ prescribers and 4+ pharmacies.(8) A national study found that 13% of patients had overlapping prescriptions from two or more different prescribers during an 18-month period. Of these, 0.5% used 4+ prescribers and 4+ pharmacies.(9) People who see multiple prescribers or use multiple pharmacies are more likely to die of drug overdoses.(6) When comparing the diagnostic odds ratio for opioid overdose events of 9 pharmacy shopping definitions, a threshold of 4 pharmacies had the highest diagnostic odds ratio.(10) Data from the California PDMP indicates that people with higher daily dosages are more likely to see multiple prescribers or go to multiple pharmacies.(11) However, there is no clear threshold at which multiple prescribers and multiple pharmacies represent lack of continuity or poorly coordinated care.

The data above suggest that efforts to prevent opioid overdose deaths should focus on strategies that target high-dose opioid users as well as persons who seek care from multiple doctors and pharmacies. This measure presently under consideration focuses on these two aspects of use of prescription opioid drugs.

References:
1. U.S. Department of Health and Human Services, Office of Disease Prevention and Health Promotion. (2014). National Action Plan

for Adverse Drug Event Prevention. Washington, DC. Accessed on: 4/9/15. Available at: http://www.health.gov/hcq/pdfs/ADE-Action-Plan-508c.pdf.
2. Liu Y, Logan JE, Paulozzi LJ, et al. Potential misuse and inappropriate prescription practices involving opioid analgesics. Am J Manag Care. 2013;19:648-65. PMID: 24304213.
3. Morasco BJ, Duckart JP, Carr TP, et al. Clinical characteristics of veterans prescribed high doses of opioid medications for chronic non-cancer pain. Pain. 2010;151:625-32. PMID: 20801580.
4. Agency Medical Directors Group (AMDG). Interagency Guideline on Opioid Dosing for Chronic Non-cancer Pain: An educational aid to improve care and safety with opioid therapy. 2010 Update. Accessed on: 4/9/15. Available at: http://www.agencymeddirectors.wa.gov/files/opioidgdline.pdf.
5. Dunn KM, Saunders KW, Rutter CM, et al. Opioid prescriptions for chronic pain and overdose: a cohort study. Ann Intern Med. 2010;152:85-92. PMID: 20083827.
6. Paulozzi LJ, Kilbourne EM, Shah NG, et al. A history of being prescribed controlled substances and risk of drug overdose death. Pain Med. 2012;13:87-95. PMID: 22026451.
7. Saunders KW, Dunn KM, Merrill JO, et al. Relationship of opioid use and dosage levels to fractures in older chronic pain patients. J Gen Intern Med. 2010;25:310-5. PMID: 20049546.
8. Katz N, Panas L, Kim M, et al. Usefulness of prescription monitoring programs for surveillance--analysis of Schedule II opioid prescription data in Massachusetts, 1996- 2006. Pharmacoepidemiol Drug Saf. 2010;19:115-23. PMID: 20014166.
9. Cepeda MS, Fife D, Chow W, et al. Assessing opioid shopping behaviour: a large cohort study from a medication dispensing database in the US. Drug Saf. 2012;35:325-34. PMID: 22339505.
10. Yang Z, Wilsey B, Bohm Michele, et. al. Defining Risk of Prescription Opioid Overdose: Pharmacy Shopping and Overlapping Prescriptions Among Long-term Opioid Users in Medicaid. The Journal of Pain. 2015;445—453. PMID 25681095.
11. Han H, Kass PH, Wilsey BL, et al. Individual and county-level factors associated with use of multiple prescribers and multiple pharmacies to obtain opioid prescriptions in California. PLoS One. 2012;7:e46246. PMID: 23049992.

S.4. **Numerator Statement:** Any member in the denominator with opioid prescription claims where the MED is greater than 120mg for 90 consecutive days or longer* AND who received opioid prescriptions from 4 or more prescribers AND 4 or more pharmacies.

*MED calculation is included in S.6 Numerator Details

S.7. **Denominator Statement:** Any member with two or more prescription claims for opioids filled on at least two separate days, for which the sum of the days supply is greater than or equal to 15.

S.10. **Denominator Exclusions:** Any member with a diagnosis for Cancer or a Prescription Drug Hierarchical Condition Category (RxHCC) 8, 9, 10, or 11 for Payment Year 2015; or RxHCC 15, 16, 17, 18, or 19 for Payment Year 2016 (see list in S.11 and S.2b); or a hospice indicator (Medicare Part D) from the enrollment database.

De.1. **Measure Type:** Process

S.23. **Data Source:** Administrative claims

S.26. **Level of Analysis:** Health Plan, Population : National, Population : State

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

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

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

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

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *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.

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 benefits for high dose opioids for chronic pain are not established and the risks for serious harms related to opioid therapy increase at higher opioid dosage. The use of multiple prescribers and pharmacies are associated with increased risks for opioid overdose. The risk for overdose increases with the number of prescribers and pharmacies.

Guidance from the Evidence Algorithm
1-No→3-No→ 7-Yes→ 8-Yes→9-Yes→Moderate

- *For process 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?*

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

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

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

- The measure was tested in three different health plan data sources – the Medicare population, one commercial heath plan, and the Medicaid population.
- The testing from the Medicare population used administrative claims data from January 1st 2013 to December 31st 2013. The Medicare rates ranged from 30.0 per 1,000 to 49.66 per 1,000. The Mean was 39.27 per 1,000 and the median was 38.7 per 1,000. The standard deviation was 8.32. The 25th percentile was 34.62 per 1,000, the 50th percentile is the median (38.70 per 1,000) and the 75th percentile was 43.35 per 1,000. The interquartile range was 8.73.
- The Medicaid rates ranged from 8.15 per 1,000 to 66.45 per 1,000. The Mean was 34.04 per 1,000 and the median was 34.29 per 1,000. The standard deviation was 20.61. The 25th percentile was 20.4 per 1,000, the 50th percentile is the median (34.29 per 1,000) and the 75th percentile was 48.1 per 1,000. The interquartile range was 27.68.
- Testing was also conducted in one Commercial health plan using administrative claims from January 1st 2013 to December 31st 2013. This plan covered 209,191 individuals age 18 and older. The measure rate for this plan was 32.03 per 1,000.

Disparities

- The beneficiary level Low Income Subsidy (LIS) variable was used to determine disparities in rates for populations with different sociodemographic status. The LIS is a subsidy paid by the Federal government to the drug plan for Medicare beneficiaries who need extra help with their prescription drug costs due to limited income and resources. The measure rate for the LIS group is 62.41 per 1,000 while the rate for the non-LIS population is significantly lower, at 28.09 per 1,000.

Questions for the Committee:

- *Specific question on information provided for gap in care.*
- *Is there a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*

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)

1. Importance to Measure and Report

1a. Evidence to Support Measure Focus

Comments: **This measure is a process measure, inherently one would not argue that being on high does opioids consecutively daily for so long (90 days) as prescribed by multiple providers would be a "risk" for complications (death, side effects) but the authors do not directly support their specific decision-tree measure on this. In fact they present some conflicting evidence also, so it is tangential. The bigger issue, not addressed are whether this represents a population more at risk of diversion or lack of treatment plan documentation

**This is a process measure using administrative claims data. Developer indicates that benefits of high dose opioids for chronic pain are not established and the risks for harm increase at higher doses. The use of multiple prescribers and pharmacies are associated with increased risk for overdose.

**The benefits for high dose opioids for chronic pain are not established and the risks for serious harms related to opioid therapy increase at higher opioid dosage. The use of multiple prescribers and pharmacies are associated with increased risks for opioid overdose. The risk for overdose increases with the number of prescribers and pharmacies.

1b. Performance Gap

Comments: **I am not sure. There is variation in rates/incidence but without knowing the patient clinical characteristics one cannot say truthfully there is a "gap" in performance. This would be different if the requirements that exist in patient care documentation were included in the analysis (to show that variability) or if the measure was to be used to audit for that type of analysis

**Measure tested in Medicare, Medicaid and one commercial plan. Rates within plans varied demonstration an opportunity for improvement, similar to the other opioid related measures. The same disparity for beneficiaries using the LIS is demonstrated in this measure.

**There is a performance gap and disparities noted in in rates among different sociodemographic statuses

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Administrative claims, health plan enrollment information

Specifications:

- This measure assesses the proportion of individuals without cancer receiving prescriptions for opioids with a daily dosage greater than 120mg morphine equivalent dose (MED) for 90 consecutive days or longer, AND who received prescriptions for opioids from four or more prescribers AND four or more pharmacies.
- The level of analysis (i.e., the measured entity) is the prescription drug health plan.
    - The developer notes that the measure also contains claims data from multiple care settings, including ambulatory, skilled nursing facility, pharmacy etc.
- The measure is stratified by the following lines of business for the health plan:
    - Commercial
    - Medicare
    - Medicaid
- The measure is reported as a rate (per 1,000 plan members).
- The measure uses health plan medical and pharmacy claims and health plan member enrollment information as its data sources.
- To identify the denominator population, the measure identifies any member with two or more prescription claims for opioids filled on at least two separate days, for which the sum of the days' supply is greater than or equal to 15.
- To derive the numerator, the measure calculates the daily MED of opioid claims for each member and identifies the days where the MED threshold (120 MEDs) is exceeded; any member for whom the MED threshold is exceeded for at least 90 consecutive days meets the criteria for the MED component of the numerator. From

this group, the measure calculates the number of unique pharmacy providers associated with an opioid prescription claim and the number of unique prescribers associated with an opioid prescription claim. Any member with four or more unique pharmacy providers AND four or more unique prescribers meets the criteria for the numerator.

- A list of opioid medications is provided in the submission form.
- The measure excludes patients with a diagnosis of cancer and patients in hospice.
- A list of administrative codes (ICD-9/10, RxHCC) identifying denominator exclusions is provided in a spreadsheet attached to the measure submission.

Questions for the Committee :

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

SUMMARY OF TESTING

Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☐ **Yes** ☐ **No**

- The developer used several data sets for reliability testing:
  - For Medicare testing, the analysis included a convenience sample of over 700 Medicare Part D prescription drug plans (comprising a total of 7,067,445 individuals aged 18 and older)
  - Testing was also conducted in one Commercial health plan (comprising a total of 209,191 individuals age 18 and older)
  - For Medicaid testing, the analysis included 8 state-based prescription drug plans covering 6 states (comprising a total of 1,437,410 individuals age 18 and older)

Method(s) of reliability testing

- To demonstrate reliability, the developer conducted a signal-to-noise analysis of the computed measure score using a beta-binomial model.
  - The developer explains that a reliability score (i.e., signal-to-noise ratio) may range from 0 to 1; a score of 0 signifies that all variation is due to measurement error (“noise”), while a score of 1 signifies that all variation represents true differences in performance scores between plans (“signal”).

Results of reliability testing

- The developer provides the results of reliability testing in a table presenting the distribution of individual plan reliability scores; the mean reliability score across all plans is **0.9208**.
- The developer suggests that a reliability score of 0.7 is the minimum threshold for reliability, and that based on the high scores achieved in the analysis, this measure should be considered reliable.

Questions for the Committee:

○ *Specific questions on the method and results of reliability testing.*
○ *Is the test sample adequate to generalize for widespread implementation?*
○ *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

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

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☐ **Yes** ☐ **Somewhat** ☐ **No**
Specification not completely consistent with evidence

Question for the Committee:
○ *Are the specifications consistent with the evidence?*

2b2. Validity testing

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.

SUMMARY OF TESTING
Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:
☒ **Face validity only**
☐ **Empirical validity testing of the measure score**

Validity testing method:

- To demonstrate validity, the developer cites their (PQA's) approach to measure development and testing.
  - This approach includes identification of important concepts by PQA member workgroups, evaluation and refinement of concepts by the PQA Quality Metrics Expert Panel (QMEP), partnership with measure development experts, and processes for review, comment, and approval by PQA members.
- The developer notes that the QMEP Panel reviewed the results of measure testing, including performance measure scores, and provided an assessment of whether measure results reflect quality of care.

Validity testing results:

- The developer reports that out of 12 QMEP members voting on the measure's face validity, 83.3 percent strongly agreed that the measure results reflected quality of care.
- In addition, the developer notes that of 95 PQA members voting on whether to endorse the measure, 72.6 voted in favor of approval.
- Five PQA member organizations also tested the measure using their own data, and all strongly agreed that the measure reflected the quality of care provided for their populations.

- ***NQF Staff Note:*** *Assessment of this measure's validity appears to have been conducted by the same groups involved in development of the measure; NQF prefers face validity to be assessed by experts or other stakeholder groups who have not been involved in the measure development process.*

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*
- *Other specific question of the validity testing?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- This measure excludes patients with a diagnosis of cancer and patients in hospice.
- The developer's rationale for these exclusions is that patients at end of life, undergoing hospice care, and those with cancer may have unusual requirements for pain management; the developer notes that these exclusions are consistent with the 2016 CDC Guideline for Prescribing Opioids for Chronic Pain.
- Because prescription claims data do not contain claims for palliative medication, such as opioids, for persons in Medicare Part D that are in hospice care, this exclusion was not tested by the developer. In addition, for the Medicaid population, the majority of the plans were not able to identify hospice exclusions in their data.
- For the cancer exclusion, the developer provided an analysis of data from eight health plans, identifying the number of exclusions and the percent of the overall population that would be affected by including patients with cancer diagnoses.
- The developer reports that the cancer patient exclusions were 0.5% to 1.9% of the overall population. The one Medicaid plan that could identify hospice exclusions found only 15 cases, which represented 0.003% of their total population.
- Interpreting the results of this analysis, the developer states that in some plans, almost 2% of the population has cancer and would be included in the measure if cancer was not excluded, suggesting that this is a significant proportion of the population that could potentially impact the measure rates.
- The developer states that no inferences about the hospice exclusion could be drawn because the majority of the plans could not identify exclusions.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

Questions for the Committee:

Do you agree with the developer that this measure does not require risk adjustment?

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- To assess the measure's ability to identify meaningful differences in performance, the developer analyzed their testing data to identify the mean, median, standard deviation, and interquartile range for the measure rates for the Medicare population and the Medicaid population.
- In addition, for the Medicaid population, the rates were divided into quartiles, and a Student's t-test was used to compare the rates between the bottom quartile (75th percentile) and top quartile (25th percentile).
- For the Medicare population, the developer reports a mean performance rate (per 1,000 members) of **3.03**, a median rate of **2.89**, and a standard deviation of **1.02**.
- For the Medicaid population, the developer reports a mean performance rate (per 1,000 members) of **2.68**, a median rate of **2.38**, and a standard deviation of **1.80**.
- The following tables provide additional results of the developer's analysis:

Table 2. Interquartile Range of Measure Rates - Medicare Population (reported as number per 1,000 members)

Minimum	1.94
25th Percentile	2.59
50th Percentile	2.89
75th Percentile	3.32
Maximum	4.41
Interquartile Range	0.73

Table 4. Interquartile Range of Measure Rates - Medicaid Population (reported as number per 1,000 members)

Minimum	0.70
25th Percentile	1.35
50th Percentile	2.38
75th Percentile	3.61
Maximum	6.12
Interquartile Range	2.26
Student's t-test p-value	0.085

- The developer's interpretation of these results is that the measure rates showed significant variation in the Medicare population, and even greater variation in the Medicaid population
- The developer also states that there is a statistically significant difference in measure rates between the top and bottom quartile of the plans included in the testing, using the less conservative alpha of 0.10 because of the small measure rates (P=0.085 at alpha=0.10), and suggests that this variation shows that there are meaningful differences in rates across plans.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

N/A

2b7. Missing Data

- The developer notes that since all data elements are available via prescription claims data, it is not expected—nor was it found—that missing data would result.
- The developer states that, as a result, performance results would not be biased, as prescription claims data provides the data elements necessary to calculate the measure rate.

Guidance from the Validity Algorithm

[Box 1] Specifications consistent with evidence → [Box 2] Potential threats to validity addressed → [Box 3] Empirical validity testing **NOT** conducted using the measure as specified → [Box 4] Face validity systematically assessed → [Box 5] Results indicate substantial agreement that performance score can be used to distinguish quality → [Moderate]

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

Committee pre-evaluation comments

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

2. Scientific Acceptability of Measure Properties

2a1. & 2b1. Specifications

Comments: **The specifications appear straightforward but they just determine an incidence or rate of occurrence. What is not

clear is whether the 4 or more providers includes or excludes "same practice" or not.
**Specifications are clear and supported by analysis/evidence.
**data elements are clear as are numerator and denominator and likely to be consistently applied

2a2. Reliability Testing

Comments: **The reliability testing is a 0.92 (good) but was performed within the same group that developed the measure. Since this is a straightforward query via claims data, I have little doubt that the reliability is there. However, if one asks more than "this is what was prescribed over time" to more like "was there a variation in appropriateness documentation" then the reliability will change. I do fin it interesting that in reading them list of developers, "pain societies" not physicians were included and could be used to test validity/reliability
**Measure score reliability testing shows a mean score across all plans of 0.9208 suggesting high reliability
**measure reliability performed mean score .9208 (min .84) . Demonstrate sufficient reliability

2b2. Validity Testing

Comments: **Reliability was on claims data in a set of multiple payers classes. However, one cannot say anything about quality with the results. Especially given the new CDC recommendations regarding controlled substances and requirements regarding ongoing care documentation, this measure may in fact be already outdated
**Face validity testing was conducted the same way as for the other opioid measures, using the developer's panel. 83.3 percent agreed that the measure reflects quality of care; 72.6 % voted in favor of approval.
**only face validity performed and as noed by NQF staff some of the experts were involved in measure development which is less than ideal. Also appears that on expert panel there is no actual provider who would prescribe this drug which may be a weakness

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **Exclusions seem appropriate. Again, se above my concerns as to how this measure does not link to quality except for 1 article. this is not to say that this is not important.
**"Exclusions are supported. No risk adjustment. Measure rates show significant variation in the Medicare population and even greater variation in the Medicaid population
**exclusion criteria is valid although developers should consider other chronic conditions such as HIV or sickle cell and have some stratification

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.

- This measure is generated or collected by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims) Other data elements include Prescription claims data.
- ALL data elements are in defined fields in electronic claims.
- Pilot test sites indicated the measure was feasible and results were able to be reported efficiently and accurately. CMS calculates the measure for Part D plans. The prescription claims and medical data is readily available.
- Certain uses of the Measures are only approved with a licensing agreement from the developer, that specifies the terms of use and the licensing fee. The developer reserves the right to determine the conditions under which it will approve and/or license the Measures.

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

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **I like the approach and this is all via claims data that is pretty straightforward. The concern I would have is how to determine 4 or more practitioners where a patient was "searching" or this was the same practice. If there was a way to link to medical care plan documentation and appropriateness of that, then this would be so much stronger
**Highly feasible based on same rationale as other opioid related measures
**no concerns about feasibility except for clincians who dont use e-prescribe or an EHR or patients who go to other states to fill prescriptions -- do all states have the same access to data

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

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

Current uses of the measure:
Publicly reported? ☐ **Yes** ☒ **No**

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

Accountability program details:

- The measure was developed in 2015.
- The measure is currently being used in the Medicare Part D Overutilization Monitoring System to monitor the utilization of opioids for members with the Medicare drug benefit.
- CMS has announced plans to move this measure into the 2019 Part D Display Measures, using data from 2017.
- Reporting of results is not yet available.

Improvement results:

- There are no improvement results, as this is the initial endorsement submission.

Potential harms:

- Although no unintended negative consequences to individuals or populations were identified during testing, , concerns have been raised that prescribing changes such as dose reduction (without offering or arranging evidence-based treatment for patients with opioid use disorder) might be associated with unintended negative consequences, such as patients seeking heroin or other illicitly obtained opioids (1,2) or interference with appropriate pain treatment.(3) Data indicate that if access to prescription opioids is limited, some users of opioid analgesics will transition to heroin or other illicitly obtained opioids, leading to increased overdose death coincident with prescribing restrictions.(1) There are also concerns about pain being underdiagnosed and undertreated, particularly for ethnic and racial minorities, females, children or infirm elderly, or individuals who may be perceived to have mental health problems or are drug seeking.(4,5)
- These concerns must be balanced by the current situation in the United States which has been described by the CDC as an epidemic of opioid abuse, overdose, and deaths. Overdose involving opioid analgesics killed almost 17,000 persons in 2010 and the number of people with opioid analgesic use disorders increased to nearly 2 million.(6) The developer believes the potential benefits of monitoring those patients receiving the very highest doses of opioids for extended periods of time or receiving these drugs from multiple providers outweighs potential negative consequences.

Feedback :

- Developer did not identify any specific feedback loops related to this 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

4. Usability and Use

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **I found myself asking the question "so who is the audience for this" - public health? healthcare plan? provider? And "what to do with the results"? Having some studies that show the link to complications or link to appropriateness of care limits the use of what could be a powerful and positive query

**Not currently publicly reported, not used in accountability programs. Planned for use by CMS in the 2019 Part D Display measures. Measure is used in Part D Overutilization Monitoring Program. Same oncerns about access to opioids as in the other opioid measures.

**not publicly reported although used in Medicare Part D Overutilization Monitoring System and CMS plans to adopt in 2019 based on the 2017 data . Also no improvement data since this is initial endorsement

Criterion 5: Related and Competing Measures

Related or competing measures

- The measure is related to 2940 and 2950 which are being proposed for endorsement.

Harmonization

- N/A

Pre-meeting public and member comments

Submitted By: ADVault, Inc.

ADVault believes that people live better lives and, if in a health crisis, can receive better care when they have confidence they can be involved in the creation and implementation of their medical treatment plans and decisions, factors extremely important when it comes to addictive, narcotic medications like opioids. To do so, they must be able to communicate and express their goals, preferences and priorities for care in a meaningful and actionable way so providers can consider those thoughts. At some point in life, everyone will lose his or her ability to communicate effectively and understand what is being asked of him or her. Healthcare agents should have the confidence to know those value statements as well, in order to fulfill their role as surrogate decision-makers. Non-surrogate family members are comforted with third-party decision-making if they have proof the patient's voice is being heard, clearly understood, and to the extent possible, honored.
Therefore, ADVault strongly recommends providers (1) search for a person's digital emergency, critical and advance care plan (ECACP) upon admission and each time the patient is transitioned to a new site of care, (2) review and update the ECACP in various stages of a person's admission (outpatient or inpatient) and/or illness to ensure respect for the person's goals, preferences and priorities for care, (3) link the digital ECACP to the EHR and/or patient portal in order to ease access and address security, privacy and patient consent concerns, (4) track and make available the number of ECACPs found, opened and re-visited, and the impact they have on the care of the patient, as well as patient, family and caregiver satisfaction, such data to be reported in a

manner such that: (a) consumers can make better choices about hospitals and doctors; (b) doctors improve the satisfaction and quality of their work; and (c) hospital administrators gauge performance and align caregiving goals with actual outcomes. Finally, if no ECACP can be found via standards-based healthcare IT transport mechanisms, the hospital/provider should engage the patient to create one whenever possible.

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Use of Opioids from Multiple Providers and at High Dosage in Persons Without Cancer
IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: 5/13/2016

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

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

Outcome

- ☐ Health outcome: Click here to name the health outcome
- ☐ Patient-reported outcome (PRO): Click here to name the PRO
  PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors
- ☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☒ Process: Prescriptions for high-doses of opioids and from multiple prescribers and pharmacies

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

It has been shown that the measured process, prescriptions for high-doses of opioids and from multiple prescribers and pharmacies, correlates with undesired health outcomes. Benefits of high-dose opioids for chronic pain are not established and the risks for serious harms related to opioid therapy increase at higher opioid dosage. Higher opioid dosages are associated with increased risks for motor vehicle injury, opioid use disorder, and overdose. The risk for overdose increases in a dose-dependent manner. Lower dosages of opioids reduce the risk for overdose, but a single dosage threshold for safe opioid use has not been identified. Use of multiple prescribers and pharmacies are associated with increased risks for opioid overdose. The risk for overdose increases with the number of prescribers and pharmacies.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?
☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***
☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***
☒ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***
☒ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION
1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

CDC Guideline: Dowell D, Haegerich TM, Chou R. CDC Guideline for Prescribing Opioids for Chronic Pain - United States, 2016. MMWR Recomm Rep. 2016 Mar 18;65(1):1-49. doi: 10.15585/mmwr.rr6501e1. Available at: http://www.cdc.gov/drugoverdose/prescribing/guideline.html.

AMDG Guideline: Interagency Guideline on Prescribing Opioids for Pain: Developed by the Washington State Agency Medical Directors' Group (AMDG) in collaboration with an Expert Advisory Panel, Actively Practicing Providers, Public Stakeholders, and Senior State Officials. June 2015. Available at: www.agencymeddirectors.wa.gov.

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

CDC Guideline: Recommendation 5, pages 22-24. "When opioids are started, clinicians should prescribe the lowest effective dosage. Clinicians should use caution when prescribing opioids at any dosage, should carefully reassess evidence of individual benefits and risks when considering increasing dosage to ≥50 morphine milligram equivalents (MME)/day, and should avoid increasing dosage to ≥90 MME/day or carefully justify a decision to titrate dosage to ≥90 MME/day (recommendation category: A, evidence type: 3)."

AMDG Guideline: Recommendation 3, page 11. "Do not escalate COAT [chronic opioid analgesic therapy] to more than 120 mg/day MED without first obtaining a consultation from a trained pain specialist who agrees that a high dose is indicated and appropriate. Providers must routinely monitor and document sustained improvement in function and quality of life and an absence of the risk factors listed in recommendations 1 and 2."

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

CDC Guideline: Category A recommendation: Applies to all persons; most patients should receive the recommended course of action.

AMDG Guideline: N/A

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

CDC Guideline: Recommendation Categories
Based on evidence type, balance between desirable and undesirable effects, values and preferences, and resource allocation (cost).
Category A recommendation: Applies to all persons; most patients should receive the recommended course of action.
Category B recommendation: Individual decision making needed; different choices will be appropriate for different patients. Clinicians help patients arrive at a decision consistent with patient values and preferences and specific clinical situations.

AMDG Guideline: N/A

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☒ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

CDC Guideline: Dowell D, Haegerich TM, Chou R. CDC Guideline for Prescribing Opioids for Chronic Pain - United States, 2016. MMWR Recomm Rep. 2016 Mar 18;65(1):1-49. doi: 10.15585/mmwr.rr6501e1. Available at: http://www.cdc.gov/drugoverdose/prescribing/guideline.html.

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

CDC Guideline: The CDC guideline was developed using the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) method (http://www.gradeworkinggroup.org). A previously published systematic review sponsored by the Agency for Healthcare Research and Quality (AHRQ) on the effectiveness and risks of long-term opioid treatment of chronic pain (http://www.effectivehealthcare.ahrq.gov/ehc/products/557/1971/chronic-pain- opioid-treatment-report-141007.pdf, http://dx.doi.org/10.7326/M14-2559) initially served to directly inform the recommendation statements. CDC conducted additional literature searches to update the AHRQ evidence review; more details about the literature search strategies and GRADE methods applied are provided in the Clinical Evidence Review

(http://stacks.cdc.gov/view/cdc/38026). CDC developed GRADE evidence tables to illustrate the quality of the evidence for each clinical question. As identified in the AHRQ-sponsored clinical evidence review, the overall evidence base for the effectiveness and risks of long-term opioid therapy is low in quality per the GRADE criteria. Thus, CDC conducted a Contextual Evidence Review (http://stacks.cdc.gov/view/cdc/38027) to provide additional information, including the epidemiology of opioid pain medication overdose. CDC constructed narrative summaries of this contextual evidence and used the information to support the clinical recommendations.

Complete section 1a.7

_____

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

CDD Guideline: The CDC Clinical Evidence Review evaluated the clinical questions regarding the effectiveness, benefits, and harms of long-term opioid therapy (use of opioids on most days for >3 months) for chronic pain. The CDC Contextual Evidence Review focused on the effectiveness of alternative treatments, benefits and harms of opioid therapy; provider and patient values and preferences; and resource allocation.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

CDC Guideline: Type 3 evidence: Observational studies or randomized clinical trials with notable limitations.

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

CDC Guideline: Evidence Type: Based on study design as well as a function of limitations in study design or implementation, imprecision of estimates, variability in findings, indirectness of evidence, publication bias, magnitude of treatment effects, dose- response gradient, and constellation of plausible biases that could change effects.
Type 1 evidence: Randomized clinical trials or overwhelming evidence from observational studies.
Type 2 evidence: Randomized clinical trials with important limitations, or exceptionally strong evidence from observational studies.
Type 3 evidence: Observational studies or randomized clinical trials with notable limitations.
Type 4 evidence: Clinical experience and observations, observational studies with important limitations, or randomized clinical trials with several major limitations.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: January 2008 through August 2014

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

CDC Guideline: In the CDC Guideline, for Key Question 2, specifically related to how harms vary depending on the opioid dose used, 6 observational studies were included from the Clinical Evidence Review (CDC p. 44, Table 1). Five additional observational studies on the association of opioid dosage and overdose risk were identified in the Contextual Evidence Review and considered in the CDC Guideline. These had been excluded from the clinical evidence review because patient samples were not restricted to patients with chronic pain only.

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

CDC Guideline: The overall quality of evidence across studies in the complete body of evidence is low in quality per the GRADE criteria. The relevant studies related to Key Question 2 were described as fair- to good-quality observational studies. These assessments were primarily related to serious study limitations. For risk of overdose related to MME/day, there was no inconsistency or imprecision, and the magnitude of effect and dose response relationship were notable.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

CDC Guideline: Evidence is insufficient (0 studies). The benefits of high-dose opioids for chronic pain are not established.

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

CDC Guideline: The Clinical Evidence Review found that risks for serious harms related to opioid therapy increase at higher opioid dosage. Higher opioid dosages are associated with increased risks for motor vehicle injury, opioid use disorder, and overdose. The Clinical and Contextual Evidence Reviews found that opioid overdose risk increases in a dose-response manner, that dosages of 50–<100 MME/day have been found to increase risks for opioid overdose by factors of 1.9 to 4.6 compared with dosages of 1–<20 MME/day, and that dosages ≥100 MME/day are associated with increased risks of overdose 2.0–8.9 times the risk at 1–<20 MME/day.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

1a.8 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.8.1 What process was used to identify the evidence?

The Centers for Medicare and Medicaid (CMS) Part D Overutilization Monitoring System (OMS) was identified on the CMS website: https://www.cms.gov/Medicare/Prescription-Drug-coverage/PrescriptionDrugCovContra/RxUtilization.html.

A PubMed search was conducted using combinations of the following search terms: opioid, overdose, doctor shopping, pharmacy shopping, multiple prescribers, multiple pharmacies. Articles referenced in the identified articles were scanned for relevance. The CDC Guideline and Clinical and Contextual Evidence Reviews were also reviewed for relevant references (CDC Guideline: Dowell D, Haegerich TM, Chou R. CDC Guideline for Prescribing Opioids for Chronic Pain - United States, 2016. MMWR Recomm Rep. 2016 Mar 18;65(1):1-49. doi: 10.15585/mmwr.rr6501e1. Available at: http://www.cdc.gov/drugoverdose/prescribing/guideline.html.; CDC Clinical Evidence Review. Available at: http://stacks.cdc.gov/view/cdc/38026; CDC Contextual Evidence Review. Available at: http://stacks.cdc.gov/view/cdc/38027).

1a.8.2. Provide the citation and summary for each piece of evidence.

1. CMS. Medicare Part D Overutilization Monitoring System (OMS) Summaryhttps://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovContra/Downloads/Fact-Sheet-Overutilization-Monitoring-System-11032015.pdf
Summary: CMS developed a comprehensive morphine equivalent dose (MED) approach to assist Part D sponsors in identifying high risk beneficiaries. Beneficiaries who are dispensed opioids that exceed 120 mg of cumulative MED for at least 90 consecutive days, and whose opioid prescriptions are associated with more than 3 prescribers and more than 3 pharmacies are identified as high-risk beneficiaries (i.e., potential opioid overutilizers). This approach was based the method used in Washington State, as well as the opioid product list and MED conversion factors maintained by the CDC. This cumulative MED approach to identify high risk use of opioids is now being widely adopted outside of Part D.

2. CMS. Advance Notice of Methodological Changes for Calendar Year (CY) 2017 for Medicare Advantage (MA) Capitation Rates, Part C and Part D Payment Policies and 2017 Call Letter. Available at: https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Downloads/Advance2017.pdf
Summary: Part D sponsors have had a significant impact on reducing overutilization of opioids and APAP. From 2011 through 2015, there was a 47% decrease or 13,753 fewer Medicare Part D beneficiaries identified as potential opioid overutilizers (i.e., beneficiaries with at least 90 consecutive days with greater than 120 mg MED daily with more than 3 [i.e., 4 or more] prescribers and more than 3 [i.e., 4 or more] pharmacies contributing to their opioid claims). This represents a 57% decrease in the share of beneficiaries using opioids who are identified as potential opioid overutilizers.

3. Franklin GM, Mai J, Turner J, Sullivan M, Wickizer T, Fulton-Kehoe D. Bending the prescription opioid dosing and mortality curves: impact of the Washington State opioid dosing guideline. *Am J Ind Med.* 2012 Apr;55(4):325-31. doi:10.1002/ajim.21998. PMID: 22213274.
Summary: In a retrospective observational study using data from WA state workers' compensation system, the 2007 introduction of an opioid dosing guideline in WA appeared to be associated temporally with a 26% decline in the average dose for long-acting opioids and a 35% decline in percent of claimants receiving opioid doses of at least 120 mg MED per day. There was a 50% decrease in opioid-related deaths among injured workers from 2009 to 2010.

4. Dowell D, Haegerich TM, Chou R. CDC Guideline for Prescribing Opioids for Chronic Pain - United States, 2016. *MMWR Recomm Rep.* 2016;65(1):1-49. doi: 10.15585/mmwr.rr6501e1. Available at: http://www.cdc.gov/drugoverdose/prescribing/guideline.html.
Summary: Recommendation 9 of the CDC Guidelines states that clinicians should review the patient's history of controlled substance prescriptions using state prescription drug monitoring program (PDMP) data to determine whether the patient is receiving opioid dosages or dangerous combinations that put him or her at high risk for overdose. Clinicians should review PDMP data when starting opioid therapy for chronic pain and periodically during opioid therapy for chronic pain, ranging from every prescription to every 3 months (recommendation category: A, evidence type: 4). The text related to recommendation 9 in the CDC guidelines states that although evidence is limited on the effectiveness of PDMP implementation at the state level on prescribing and mortality outcomes, most fatal overdoses have been shown to be associated with patients receiving opioids from multiple prescribers and/or with patients receiving high total daily opioid dosages; and information on both of these risk factors for overdose are available to prescribers in the prescription drug monitoring program (PDMP)." "Clinicians should discuss safety concerns, including increased risk for respiratory depression and overdose, with patients found to be receiving opioids from more than one prescriber or receiving medications that increase risk when combined with opioids (e.g., benzodiazepines) and consider offering naloxone (see Recommendation 8)."

5. Yang Z, Wilsey B, Bohm M, Weyrich M, Roy K, Ritley D, Jones C, Melnikow J. Defining risk of prescription opioid overdose: pharmacy shopping and overlapping prescriptions among long-term opioid users in medicaid. *J Pain.* 2015;16(5):445-53. doi: 10.1016/j.jpain.2015.01.475. PubMed PMID: 25681095.
Summary: An analysis of multistate Medicaid claims database (2008-2010) was conducted to evaluate strategies for identifying patients at high risk for overdose among enrollees who used 3 or more opioid prescriptions for 90 or

more days. Diagnostic odds ratios were compared the for opioid overdose events of 9 pharmacy shopping definitions. The diagnostic odds ratio for the criterion of 4 or more pharmacies in a 90-day period had the highest value at 5.40. The percentage of patients with opioid overdose events increased as the number of pharmacies increased.

6. Gwira Baumblatt JA, Wiedeman C, Dunn JR, Schaffner W, Paulozzi LJ, Jones TF. High-risk use by patients prescribed opioids for pain and its role in overdose deaths. *JAMA Intern Med* 2014;174(5):796–801. PMID: 24589873.
Summary: A matched case-control study among patients prescribed opioids in Tennessee (2008-2011) found an increased risk of opioid-related overdose death with 4 or more prescribers (adjusted odds ratio [aOR], 6.5; 95% Confidence Interval [CI], 5.1-8.5), 4 or more pharmacies (aOR, 6.0; 95% CI, 4.4-8.3), and more than 100 morphine milligram equivalents (MMEs) per day (aOR, 11.2; 95% CI, 8.3-15.1) daily mean dose. At least one of these risk factors was present in 55% of overdose deaths. Risk of overdose death increased with increasing number of pharmacies ($P < .001$) and prescribers used by the patient ($P < .001$).

7. Cepeda MS, Fife D, Chow W, Mastrogiovanni G, Henderson SC. Assessing opioid shopping behaviour: a large cohort study from a medication dispensing database in the US. *Drug Saf.* 2012;35(4):325-34. doi: 10.2165/11596600-000000000-00000. PMID: 22339505.
Summary: A cohort study of prescription data was conducted to provide a definition of shopping behavior that differentiates opioids from benzodiazepines and diuretics, avoiding the inappropriate flagging of individuals with legitimate use of opioids. The authors concluded that having 2 or more overlapping prescriptions written by different prescribers and filled at 3 or more pharmacies differentiates opioids from diuretics and likely constitutes shopping behavior.

8. Paulozzi LJ, Kilbourne EM, Shah NG, Nolte KB, Desai HA, Landen MG, Harvey W,Loring LD. A history of being prescribed controlled substances and risk of drug overdose death. *Pain Med.* 2012;13(1):87-95. doi: 10.1111/j.1526-4637.2011.01260.x. PMID: 22026451.
Summary: A matched case-control study in New Mexico (2006-2008) showed that risk of unintentional overdose death increased with the number of prescribers and pharmacies. The odds ratio of one more prescriber and pharmacy in the previous six months was 1.7 (95% CI, 1.6-1.9) and 2.3 (95% CI, 2.0-2.5), respectively.

9. Peirce GL, Smith MJ, Abate MA, Halverson J. Doctor and pharmacy shopping for controlled substances. *Med Care.* 2012;50(6):494-500. doi: 10.1097/MLR.0b013e31824ebd81. PubMed PMID: 22410408.
Summary: In a case-control study in West Virginia (2005-2007), subjects classified as doctor shoppers (4 or more prescribers in the previous 6 months) had 2 times the odds (OR, 2.0; 95% CI, 1.6–2.6), and pharmacy shoppers (4 or more pharmacies in the previous 6 months) had 3 times the odds (3.2; 95% CI, 2.3–4.5) of drug-related death compared with those classified as non-shoppers. Subjects classified as both doctor and pharmacy shoppers also had increased odds (OR, 3.6; 95% CI, 2.7–4.7) of drug-related death compared with non-shoppers.

10. Gomes T, Mamdani MM, Dhalla IA, Paterson JM, Juurlink DN. Opioid dose and drug-related mortality in patients with nonmalignant pain. *Arch Intern Med.* 2011;171(7):686-91. doi: 10.1001/archinternmed.2011.117. PubMed PMID: 21482846.
Summary: In an observational nested case-control study of decedents in Ontario Canada (1997-2006) who received an opioid prescription, those whose deaths were related to opioids were more likely to have obtained opioids from multiple physicians and pharmacies than decedents whose deaths were not related to opioids.

11. Katz N, Panas L, Kim M, Audet AD, Bilansky A, Eadie J, Kreiner P, Paillard FC, Thomas C, Carrow G. Usefulness of prescription monitoring programs for surveillance--analysis of Schedule II opioid prescription data in Massachusetts, 1996-2006. *Pharmacoepidemiol Drug Saf.* 2010;19(2):115-23. doi:10.1002/pds.1878. PMID: 20014166.
Summary: An analysis of Massachusetts prescription monitoring program (PMP) data was conducted to evaluate trends in opioid prescribing, dispensing, and usage of prescription data. The authors selected the criterion of 4 or more prescribers and 4 or more pharmacies as an indicator of potential non-medical use and diversion of prescription opioids such as doctor shopping. The authors commented that the criterion of 3 or prescribers and 3 or

more pharmacies is not stringent enough and likely to misclassify patients who are using opioids appropriately (false positives).

12. White AG, Birnbaum HG, Schiller M, Tang J, Katz NP. Analytic models to identify patients at risk for prescription opioid abuse. *Am J Manag Care.* 2009;15(12):897-906. PubMed PMID: 20001171. Available at: http://www.ajmc.com/journals/issue/2009/2009-12-vol15-n12/AJMC_09Dec_White_897to906/
Summary: A retrospective observational study of privately insured patients in Maine (2005-2006) showed an increased risk for opioid dependence, abuse, or overdose among persons receiving multiple prescriptions, having multiple prescribers, or using multiple pharmacies.

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
3_PQA-Opioids_High_Dose_Multi_Provider_Evidence_Form_051016.docx

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., the benefits or improvements in quality envisioned by use of this measure)*
Abuse and overdose of prescription drugs is a major public health issue in the United States.(1,2) Though there is no FDA maximum dose or duration for opioid drugs, studies have demonstrated that patient populations taking high opioid doses for prolonged periods are often characterized by high rates of psychiatric and substance abuse disorders, frequently do not receive care consistent with clinical guidelines, and have higher death rates.(3-6) Studies have shown that people who see multiple prescribers or use multiple pharmacies are more likely to die of drug overdoses.(6)

PQA developed 3 measures related to prescription opioid use that are indicative of the quality of care for patients taking these medications. The measures examine the quality of use related to the dose of the medications over time, and access to the medications through multiple providers. This measure, Use of Opioids from Multiple Providers in Persons Without Cancer, focuses on the use of opioids at high dose and from multiple providers.

Claims data from commercially insured patients indicate that approximately 8% of opioid prescriptions for acute pain and 12% for chronic pain specify a daily dosage of 120mg MED or more.(2) The Washington State Agency Medical Directors Group has suggested 120mg MED as a dosage level that should not be exceeded without special consideration.(4) Group Health Cooperative (GHC), which implemented this guidance from the 2010 edition, has demonstrated a reduction in opioid doses for their patients with chronic pain. For the last quarter of 2014, less than one-quarter of these patients seen by GHC providers received 50 mg/day MED or greater and only 7.3% exceeded 120 mg/day MED.(4) The proportion of patients being treated at this dosage for more than 90 days has not been described. However, one study of veterans treated with 180mg MED/day or more for 90+ days (3) found that this group was characterized by high rates of psychiatric and substance abuse disorders and frequently did not receive care consistent with clinical guidelines. Studies suggest that high opioid dosage increases the risk of overdoses and fractures.(5-7)

Prescription drug monitoring programs (PDMP), which track the use of multiple providers by patients, indicate that such use is typically found among a small proportion of patients, with the proportion declining as the number of providers increases. In Massachusetts in 2006, considering only Schedule II opioids, 0.5% of patients saw 4+ prescribers and 4+ pharmacies.(8) A national study found that 13% of patients had overlapping prescriptions from two or more different prescribers during an 18-month period. Of these, 0.5% used 4+ prescribers and 4+ pharmacies.(9) People who see multiple prescribers or use multiple pharmacies are more likely to die of drug overdoses.(6) When comparing the diagnostic odds ratio for opioid overdose events of 9 pharmacy shopping definitions, a threshold of 4 pharmacies had the highest diagnostic odds ratio.(10) Data from the California PDMP indicates that people with higher daily dosages are more likely to see multiple prescribers or go to multiple pharmacies.(11) However, there is no clear threshold at which multiple prescribers and multiple pharmacies represent lack of continuity or poorly coordinated care.

The data above suggest that efforts to prevent opioid overdose deaths should focus on strategies that target high-dose opioid users as well as persons who seek care from multiple doctors and pharmacies. This measure presently under consideration focuses on these two aspects of use of prescription opioid drugs.

References:

1. U.S. Department of Health and Human Services, Office of Disease Prevention and Health Promotion. (2014). National Action Plan for Adverse Drug Event Prevention. Washington, DC. Accessed on: 4/9/15. Available at: http://www.health.gov/hcq/pdfs/ADE-Action-Plan-508c.pdf.
2. Liu Y, Logan JE, Paulozzi LJ, et al. Potential misuse and inappropriate prescription practices involving opioid analgesics. Am J Manag Care. 2013;19:648-65. PMID: 24304213.
3. Morasco BJ, Duckart JP, Carr TP, et al. Clinical characteristics of veterans prescribed high doses of opioid medications for chronic non-cancer pain. Pain. 2010;151:625-32. PMID: 20801580.
4. Agency Medical Directors Group (AMDG). Interagency Guideline on Opioid Dosing for Chronic Non-cancer Pain: An educational aid to improve care and safety with opioid therapy. 2010 Update. Accessed on: 4/9/15. Available at: http://www.agencymeddirectors.wa.gov/files/opioidgdline.pdf.
5. Dunn KM, Saunders KW, Rutter CM, et al. Opioid prescriptions for chronic pain and overdose: a cohort study. Ann Intern Med. 2010;152:85-92. PMID: 20083827.
6. Paulozzi LJ, Kilbourne EM, Shah NG, et al. A history of being prescribed controlled substances and risk of drug overdose death. Pain Med. 2012;13:87-95. PMID: 22026451.
7. Saunders KW, Dunn KM, Merrill JO, et al. Relationship of opioid use and dosage levels to fractures in older chronic pain patients. J Gen Intern Med. 2010;25:310-5. PMID: 20049546.
8. Katz N, Panas L, Kim M, et al. Usefulness of prescription monitoring programs for surveillance--analysis of Schedule II opioid prescription data in Massachusetts, 1996- 2006. Pharmacoepidemiol Drug Saf. 2010;19:115-23. PMID: 20014166.
9. Cepeda MS, Fife D, Chow W, et al. Assessing opioid shopping behaviour: a large cohort study from a medication dispensing database in the US. Drug Saf. 2012;35:325-34. PMID: 22339505.
10. Yang Z, Wilsey B, Bohm Michele, et. al. Defining Risk of Prescription Opioid Overdose: Pharmacy Shopping and Overlapping Prescriptions Among Long-term Opioid Users in Medicaid. The Journal of Pain. 2015;445—453. PMID 25681095.
11. Han H, Kass PH, Wilsey BL, et al. Individual and county-level factors associated with use of multiple prescribers and multiple pharmacies to obtain opioid prescriptions in California. PLoS One. 2012;7:e46246. PMID: 23049992.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

The measure was tested in three different health plan data sources – the Medicare population, one commercial heath plan, and the Medicaid population.

The testing from the Medicare population used administrative claims data from January 1st 2013 to December 31st 2013. For the Medicare testing, the analysis included a convenience sample of over 700 Medicare Part D prescription drug plans, covering 7,067,445 individuals aged 18 and older.

The Medicare rates ranged from 1.94 per 1,000 to 4.41 per 1,000. The mean was 3.03 per 1,000 and the median was 2.89 per 1,000. The standard deviation was 1.02. The 25th percentile was 2.59 per 1,000, the 50th percentile is the median (2.89 per 1,000) and the 75th percentile was 3.32 per 1,000. The interquartile range was 0.73

The majority of testing used Medicaid prescription claims data from January 1st 2015-December 31st 2015. Testing also included prescription claims data from one state's Medicaid plan from July 1st 2014-June 30th 2015. Testing included 8 state based prescription drug plans in 6 states, covering 1,437,410 individuals age 18 and older.

The Medicaid rates ranged from 0.70 per 1,000 to 6.12 per 1,000. The mean was 2.68 per 1,000 and the median was 2.38 per 1,000. The standard deviation was 1.80. The 25th percentile was 1.35 per 1,000, the 50th percentile is the median (2.38 per 1,000) and the 75th percentile was 3.61 per 1,000. The interquartile range was 2.26.

Testing was also conducted in one Commercial health plan using administrative claims from January 1st 2013 to December 31st 2013. This plan covered 209,191 individuals age 18 and older. The measure rate for this plan was 1.45 per 1,000.

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*
Disparities data is available for the Medicare population. The testing from the Medicare population used administrative claims data from January 1st 2013 to December 31st 2013 and included a convenience sample of over 700 Medicare Part D prescription drug plans, covering 7,067,445 individuals aged 18 and older.

The beneficiary level Low Income Subsidy (LIS) variable was used to determine disparities in rates for populations with different sociodemographic status. The LIS is a subsidy paid by the Federal government to the drug plan for Medicare beneficiaries who need extra help with their prescription drug costs due to limited income and resources. The measure rate for the LIS group is 6.48 per 1,000 while the rate for the non-LIS population is significantly lower, at 1.41 per 1,000.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
N/A

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, A leading cause of morbidity/mortality, Patient/societal consequences of poor quality
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
The misuse of prescription opioids in America is a public health crisis and addressing the overdose epidemic is a high priority for the US government.(1-4) Deaths from drug overdose have risen steadily over the past two decades and have become the leading cause of injury death in the United States.(5) Since 1999, prescription opioid use and overdose deaths have quadrupled.(6) More than 165,000 people have died from prescription opioids in this timeframe,(7) yet there has not been an overall change in the amount of pain that Americans report.(8,9) In 2014, more than 14,000 people died from prescription opioid overdose, more than any year on record.(7) Higher opioid dosages are associated with increased risks for motor vehicle injury, opioid use disorder, and overdose.(2) The risk for overdose increases in a dose-dependent manner and lower dosages of opioids reduce the risk for overdose.(2) Use of multiple prescribers and pharmacies are associated with increased risks for opioid overdose and the risk for overdose increases with the number of prescribers and pharmacies.(10-13) Identifying patients at higher risk for overdose is a component of improving opioid prescribing practices to reduce the risk of overdose.(2)

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
1. S.524 - Comprehensive Addiction and Recovery Act of 2016. Available at: https://www.congress.gov/bill/114th-congress/senate-bill/524/text.
2. Dowell D, Haegerich TM, Chou R. CDC Guideline for Prescribing Opioids for Chronic Pain - United States, 2016. MMWR Recomm Rep. 2016;65:1-49. doi:10.15585/mmwr.rr6501e1. (PMID: 26987082).
3. HHS. ASPE Issue Brief: Opioid Abuse in the U.S. and HHS Actions to Address Opioid-Drug Related Overdoses and Deaths; 2015. Available at: http://aspe.hhs.gov/basic-report/opioid-abuse-us-and-hhs-actions-address-opioid-drug-related-overdoses-and-deaths.
4. US Department of Health and Human Services. National Action Plan for Adverse Drug Event Prevention. Washington, DC; 2014. Available at: http://health.gov/hcq/ade.asp.

5. CDC. Web-based Injury Statistics Query and Reporting System (WISQARS). 2014. Available at: http://www.cdc.gov/injury/wisqars/fatal.html.
6. Chen LH, Hedegaard H, Warner M. Drug-poisoning Deaths Involving Opioid Analgesics: United States, 1999-2011. NCHS Data Brief. 2014;(166):1-8
7. CDC. Wide-ranging online data for epidemiologic research (WONDER). Atlanta, GA: CDC, National Center for Health Statistics; 2016. Available at: http://wonder.cdc.gov.
8. Chang H, Daubresse M, Kruszewski S, et al. Prevalence and treatment of pain in emergency departments in the United States, 2000 – 2010. Amer J of Emergency Med 2014; 32(5): 421-31.
9. Daubresse M, Chang H, Yu Y, Viswanathan S, et al. Ambulatory diagnosis and treatment of nonmalignant pain in the United States, 2000 – 2010. Medical Care 2013; 51(10): 870-878.
10. Gwira Baumblatt JA, Wiedeman C, Dunn JR, Schaffner W, Paulozzi LJ, Jones TF. High-risk use by patients prescribed opioids for pain and its role in overdose deaths. JAMA Intern Med 2014;174(5):796–801. (PMID: 24589873.) ?
11. Paulozzi LJ, Kilbourne EM, Shah NG, Nolte KB, Desai HA, Landen MG, Harvey W,Loring LD. A history of being prescribed controlled substances and risk of drug overdose death. Pain Med. 2012;13(1):87-95. doi: 10.1111/j.1526-4637.2011.01260.x. PMID: 22026451.
12. Peirce GL, Smith MJ, Abate MA, Halverson J. Doctor and pharmacy shopping for controlled substances. Med Care. 2012;50(6):494-500. doi: 10.1097/MLR.0b013e31824ebd81. PubMed PMID: 22410408.
13. Gomes T, Mamdani MM, Dhalla IA, Paterson JM, Juurlink DN. Opioid dose and drug-related mortality in patients with nonmalignant pain. Arch Intern Med. 2011;171(7):686-91. doi: 10.1001/archinternmed.2011.117. PubMed PMID: 21482846.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
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 subcriteria 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):*
Behavioral Health : Alcohol, Substance Use/Abuse, Mental Health : Alcohol, Substance Use/Abuse

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Safety : Medication Safety

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.)*
http://pqaalliance.org/measures/default.asp

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

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** Cancer_Exclusion_RxHCC-_ICD-9_and_10_Codes-635969265833553126.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
N/A

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population,*

i.e., cases from the target population with the target process, condition, event, or outcome)
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Any member in the denominator with opioid prescription claims where the MED is greater than 120mg for 90 consecutive days or longer* AND who received opioid prescriptions from 4 or more prescribers AND 4 or more pharmacies.

*MED calculation is included in S.6 Numerator Details

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
The measurement year.

S.6. **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, 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.
Any member in the denominator with opioid prescription claims greater than 120mg MED for 90 consecutive days or longer* AND who received opioid prescriptions from 4 or more prescribers AND 4 or more pharmacies(See Table Opioids-A: Opioid Medications)

*Identifying members with prescription opioids that exceeded the MED threshold:
To identify members with prescription opioids that exceeded the MED threshold, each claim is to be converted into the MED using the appropriate conversion factor associated with the opioid product of that prescription claim (see Appendix A). The MED for each day's claims then are summed to determine the total MED for that day.
For each member in the denominator:
1. Calculate the MED for each opioid prescription claim during the measurement period, using the following equations:
• # of Opioid Dosage Units per day = (Opioid claim quantity) / (Opioid claim days supply)
• MED Daily Dose per claim = (# of opioid dosage units per day) X (# mg opioid per dosage unit) X (MED conversion factor)
2. Sum the daily MEDs of all opioid claims for each day to arrive at a total daily MED for each member.
3. Identify the days where the MED threshold is exceeded.
4. Any member, for whom the MED threshold is exceeded for 90 consecutive days or longer, meets the criteria for the MED component of the numerator.
5. From the members meeting the criteria for the MED component of the numerator (4), calculate the number of unique pharmacy providers associated with an opioid prescription claim.
6. From the members meeting the criteria for the MED component of the numerator (4), calculate the number of unique prescribers associated with an opioid prescription claim.
7. From the members meeting the criteria for the MED component of the numerator (4), any member with four or more unique pharmacy providers AND four or more unique prescribers meets the criteria for the Numerator.

Table Opioid-A: Opioid Medications (MED conversion factor)
buprenorphine patch (12.6) buprenorphine tab or film (10) butorphanol (7) codeine (0.15) dihydrocodeine (0.25)
fentanyl buccal or SL tablets, or lozenze/troche (0.13) fentanyl film or oral spray (0.18) fentanyl nasal spray (0.16)
fentanyl patch (7.2) hydrocodone (1) hydromorphone (4) levorphanol (11) meperidine (0.1)
methadone (3) morphine (1) opium (1) oxycodone (1.5) oxymorphone (3)
pentazocine (0.37) tapentadol (0.4) tramadol (0.1)

*Note: Injectables and Opioid cough and cold products and combination products containing buprenorphine and naloxone (e.g., BunavailTM, Suboxone®, Zubsolv®) are excluded from the MED calculations. Ionsys® (fentanyl transdermal patch) is also excluded as it is only for inpatient use; It is also only available through a restricted program under a Risk Evaluation and Mitigation Strategy (REMS)

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Any member with two or more prescription claims for opioids filled on at least two separate days, for which the sum of the days supply is greater than or equal to 15.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, Populations at Risk : Dual eligible beneficiaries

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
Any member with two or more prescription claims for opioids filled on at least two separate days, for which the sum of the days supply is greater than or equal to 15.

Table Opioid-A: Opioid Medications

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

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Any member with a diagnosis for Cancer or a Prescription Drug Hierarchical Condition Category (RxHCC) 8, 9, 10, or 11 for Payment Year 2015; or RxHCC 15, 16, 17, 18, or 19 for Payment Year 2016 (see list in S.11 and S.2b); or a hospice indicator (Medicare Part D) from the enrollment database.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Hospice exclusion: Exclude those members identified in the Medicare Enrollment Database as being enrolled in hospice.

Cancer exclusion: For Payment Year 2015: RxHCC 8, 9, 10, or 11. For Payment Year 2016: RxHCC 15, 16, 17, 18, or 19
ICD 9 and 10 Codes to Identify Cancer: Please see attachment in S2.b

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
The measure is stratified by the following lines of business for the health plan:

- Commercial
- Medicare
- Medicaid

Medicare Plans are further stratified by Low Income Subsidy status
Definition: Medicare Low Income Subsidy (LIS)
A subsidy paid by the Federal government to the drug plan for Medicare beneficiaries who need extra help with their prescription drug costs due to limited income and resources. Medicare beneficiaries apply for the LIS with the Social Security Administration or their State Medicaid agency.

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

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
N/A

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
N/A

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **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.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
Step One:
Calculate the denominator by identifying the number of all eligible members with two or more prescription claims for opioids filled on at least two separate days, for which the sum of the days supply is greater than or equal to 15.

Step Two:
Calculate the numerator by:
For each member in the denominator:
a. Calculate the MED for each opioid prescription claim during the measurement period, using the following equations:
• # of Opioid Dosage Units per day = (Opioid claim quantity) / (Opioid claim days supply)
• MED Daily Dose per claim = (# of opioid dosage units per day) X (# mg opioid per dosage unit) X (MED conversion factor)

b. Sum the daily MEDs of all opioid claims for each day to arrive at a total daily MED for each member.

c. Identify the days where the MED threshold is exceeded.

d. Any member, for whom the MED threshold is exceeded for 90 consecutive days or longer, meets the criteria for the MED component of the numerator.

Step Three: From those members meeting the MED component in (Step 2d.) identify those members who received opioids from 4 or more prescribers AND 4 or more pharmacies.
a. Calculate the number of unique pharmacy providers associated with an opioid prescription claim.
b. Calculate the number of unique prescribers associated with an opioid prescription claim.
c. Any member from Step 2d with four or more unique pharmacy providers AND four or more unique prescribers meets the criteria for the Numerator.

Step Four:
Divide the number of members that met the criteria in numerator (Step Three c.) by the denominator (Step One) and multiply times 1000. The rate is reported as a proportion: XX out of 1,000 members.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
N/A

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
N/A

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Health Plan Medical and Pharmacy Claims. Health Plan member enrollment information.

S.25. **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.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Health Plan, Population : National, Population : State

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Other, Pharmacy
If other: The level of analysis for this measure is the prescription drug health plan, but it contains claims data from multiple care settings, including ambulatory, skilled nursing facility, pharmacy etc.

S.28. **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

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
PQA_High_Dose_Multiprovider_testing_attachment.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

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

Measure Title Use of Opioids from Multiple Providers and at High Dosage in Persons Without Cancer

Date of Submission: 5/13/2016

Type of Measure:

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

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]
AND
If patient preference (e.g., informed decision making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;
OR
there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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.23*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ administrative claims	☒ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

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

The measure was tested in three different health plan data sources – the Medicare population, one commercial heath plan, and the Medicaid population.

For the Medicare population, data used for testing came from three different sources. The Medicare Part D Prescription Drug Event (PDE) claims were used for the identification of prescription drugs and cancer exclusions. To identify dates of birth and continuous enrollment, the Common Medicare Environment (CME) data source was used. To identify hospice enrollment, the Medicare Enrollment Database (EDB) was used.

The data source for the Commercial population came from the health plans' enrollment data, medical claims, and prescription claims.

For the Medicaid population, the data used for testing came from Medicaid administrative claims. Six Medicaid plans covering four states were included in the testing using data from a Pharmacy Benefits Manager (PBM) organization. In addition, two other state-based plans were included in the testing using their state Medicaid administrative claims database. Medical claims were used to identify the cancer diagnoses, and the pharmacy claims were used for the identification of prescription drugs.

1.3. What are the dates of the data used in testing? Click here to enter date range

The testing from the Medicare and Commercial populations used administrative claims data from January 1st 2013 to December 31st 2013. The majority of testing used Medicaid prescription claims data from January 1st 2015-Decemer 31st 2015. The data from this time period were the most complete recent data available at the time of testing. Testing also included prescription claims data from one state's Medicaid plan from July 1st 2014-June 30th 2015.

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.26*)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☒ health plan	☒ health plan
☐ other: Click here to describe	☐ other: Click here to describe

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

For the Medicare testing, the analysis included a convenience sample of over 700 Medicare Part D prescription drug plans.

Testing was also conducted in one Commercial health plan. The size and characteristics of these populations are included at the patient level in 1.6.

For the Medicaid testing, the analysis included 8 state based prescription drug plans covering 6 states. 3 plans were from the same state in the Mid-Atlantic region of the United States (US), 2 plans were from states in the South Atlantic region of the US, two plans were from states in the West South Central region of the US, and one plan was from a state

in the East South Central region of the US. The size and characteristics of the population are included at the patient level in 1.6.

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

For the Medicare testing, a total of 7,067,445 individuals aged 18 and older were included in the testing and analysis. This data can be stratified by age, gender, and type of Part D plan. Of all persons, 2,531,712 (35.8%) are male, and 4,535,732 (64.2%) are female. Individuals by age group included 271,635 (3.8%) age 18-40, 2,159,384 (30.6%) age 41-64 and 4,636,425 (65.6%) over age 65. Of all individuals, 2,492,658 (35.3%) are enrolled in a Medicare Advantage Prescription Drug Plan (MA-PD) and 4,574,787 (64.7%) are enrolled in a standalone Prescription Drug Plan (PDP).

For the Commercial plan, a total of 209,191 individuals age 18 and older were included in the analysis. Of all persons 92,227 (44.1%) are male, and 116,964 (55.9%) are female. Persons by age group included 46,913 (22.4%) age 18-40, 133,207 (63.7%) age 40-64 years, and 29,071 (13.9%) age 65 and older.

For the Medicaid plans, a total of 1,437,410 individuals age 18 and older were included in the analysis. Of all persons 515,164 (35.8%) are male, and 922,246 (64.2%) are female. Persons by age group included 897,641 (62.4%) age 18-40, 454,528 (31.6%) age 40-64 years, and 85,241 (6.0%) age 65 and older.

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

The Medicaid data was used to test reliability. This data does not include the RxHCC indicator to identify cancer exclusions, and instead uses ICD-9 or ICD-10 (depending on the year of the data) to identify diagnostic criteria for the cancer exclusions. The Medicaid data also does not allow for identification of hospice patients.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

For the Medicare population, the beneficiary level Low Income Subsidy (LIS) variable was used to determine disparities in rates for populations with different sociodemographic status. The LIS is a subsidy paid by the Federal government to the drug plan for Medicare beneficiaries who need extra help with their prescription drug costs due to limited income and resources. For the Commercial and Medicaid other populations, no patient level indicators of sociodemographic status were available in the data.

2a2. RELIABILITY TESTING

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

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

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

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

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

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

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

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

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

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

Using the parameter estimates from the Beta-Binomial model we computed individual plan reliability scores. Table 1 below shows the distribution of the plan-level scores. Plans have very high reliability scores. The reliability score mean is 0.9208 and the median 0.9337.

Table 1. Individual Plan Reliability Score Distribution

Statistic	Values
Mean	0.9208
Standard Dev.	0.0528
Min	0.8408
p10	0.8540
p25	0.8843
p50 (Median)	0.9337
p75	0.9545
p90	0.9697
max	0.9932

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

The reliability score mean is 0.9208 and the median 0.9337. A reliability score of 0.7 is the minimum threshold for reliability. Based on the high reliability scores for each of the plans in the analysis, the measure is considered reliable.

2b2. VALIDITY TESTING

2b2.1. 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*)

2b2.2. 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)*

PQA uses a systematic, transparent, consensus-based measure development and testing process. That process used in 2014 to develop this measure is outlined below:

- Step 1: PQA workgroups identify measure concepts that may be appropriate for development into fully specified performance measures. The workgroups focus on specific aspects of the medication-use system and/or specific therapeutic areas. The workgroups are open to all members of PQA and use a consensus-based approach to identify, prioritize and recommend the measure concepts that are deemed to be highly important for supporting quality improvement related to medications.

- Step 2: The measure concepts that are recommended for further development through a vote by the PQA workgroups are forwarded to the PQA Quality Metrics Expert Panel (QMEP) for evaluation and refinement. The QMEP is composed of PQA members who have backgrounds in pharmacy, medicine, research, quality improvement and measures development. The QMEP reviews the measure concepts to provide an initial assessment of the key properties of performance measures (i.e., feasibility, usability and scientific validity). The measure concepts that are rated highly on these key properties will then undergo technical specification.

- Step 3: The draft measure is provided to PQA member organizations for their comments prior to preparing technical specifications for pilot testing. The QMEP reviews member comments, edits the draft measure accordingly and poses testing questions based on this all-member feedback.

- Step 4: PQA selects partners to test the draft measure. These partners are often PQA member health plans or academic institutions with expertise in quality and performance measure testing. The testing partner implements the draft technical specifications with their existing datasets and provides a report to PQA that details testing results and recommendations for modifications of the technical specifications.

- Step 5: The workgroup that developed the measure reviews the testing results and provides comment. The QMEP reviews the workgroup comments, testing results, recommendations and potential modifications and provides a final assessment of the feasibility and scientific validity of the draft performance measures.

- Step 6: Measures that are recommended by the QMEP for endorsement are posted on the PQA web site for member review, written comments are requested, and a conference call for member organizations is scheduled to address any questions. This process allows members to discuss their views on the measures in advance of the voting period.

- Step 7: PQA member organizations, which include organizations such as large pharmacy chains, health plans, quality organizations and pharmaceutical companies vote on the performance measure(s) considered for approval and/or endorsement.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

The measure was assessed for face validity (i.e., whether it appears to measure what it intends to measure) through review by the PQA workgroup that developed the measure, the PQA Quality Metrics Expert Panel (QMEP), and PQA's full membership. In addition, feedback about validity of the measure was sought out by the five PQA member organizations who tested the measure using their own data.

The PQA Medication Use Safety Workgroup was composed of 72 PQA members that worked on multiple measure concepts. After the workgroup completed the development of the measure specifications, 37 members of the workgroup voted to determine if the draft measure should continue on further development and review by the PQA QMEP. 94.6% of members recommended that the measure move on for QMEP review.

The PQA QMEP is a panel that includes individuals with expertise and experience in pharmacy, medicine, research, and clinical or other technical expertise related to quality improvement and measure development. The names and credentials of the QMEP Panel are listed in Table 1. The QMEP reviewed the measure prior to testing to ensure scientific soundness and usefulness. The QMEP reviewed the results of the measure testing including the performance measure scores reported by plan referenced in Section 2b5 (below). Out of the 12 members on the QMEP who voted, 83.3% strongly agreed that the measure results reflected the quality of care, and recommended that the measure be considered for endorsement by the PQA membership.

Table 1. **PQA Quality Metrics Expert Panel (QMEP)**

QMEP Member Name and Credentials	QMEP Member Organization
Bimal Patel, Pharm D, MS	MedImpact
Catherine Coast, PharmD	Highmark
Chris DuPaul, MBA	CVS Caremark
Christopher Dezii, RN, MBA, CPHQ	Bristol-Myers Squibb
Christopher Powers, PharmD	CMS
David Nau, PhD, RPh, CPHQ	Pharmacy Quality Solutions
Gary Erwin, PharmD	OmniCare
Gary Young, JD, PhD	Northeastern University
Jenny Weber, PharmD, MS, PCPS,CGP, BCACP	Humana
Jessica Frank, PharmD	OutcomesMTM
Karen Farris, PhD	University of Michigan
Keith Widmer, RPh, BCPP	Express Scripts
Kent Summers, RPh, PhD	Astellas
Lynn Deguzman, PharmD, CGP	Kaiser Permanente

QMEP Member Name and Credentials	QMEP Member Organization
Mary Ann Kliethermes, PharmD	Midwestern University
Mitzi Wasik, PharmD, PCPS	Coventry Health Care/Aetna
Pat Gleason, Pharm D, BCPS	Prime Therapeutics
Steve Riddle, PharmD, BCPS	Wolters Kluwer Health
Steven Burch, RPh, PhD	GlaxoSmithKline
Tony Willoughby, PharmD	HealthMart-McKesson

PQA membership was notified prior to the PQA Annual Meeting in May 2015, of the opportunity to consider and vote for the performance measure during the meeting. (Note: PQA membership comprises health plans, community pharmacy, long-term care pharmacies, HIT companies, PBMs, healthcare quality and standards organizations, professional and trade associations, and others.) Members received the measure description, key points and evidence, measure specifications, and the performance measure scores reported by plan. During the PQA Business meeting, the measure was reviewed. Nearly all of PQA membership had a representative at the Annual Meeting and were present for the vote. Voting options included, "Agree" (indicating that the organization approved the measure), "Disagree (indicating that the organization opposed the measure) and "Abstain." Out of the 95 number of PQA members who participated in voting, 72.6% of the membership voted in favor of endorsing the measure.

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

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

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

_____

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

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

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

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

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

2b3.2. 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*)

Of the eight health plans included in the analysis, the cancer patient exclusions were 0.5% to 1.9% of the overall population. The one Medicaid plan that could identify hospice exclusions found only 15 cases, which represented 0.003% of their total population.

2b3.3. 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 results show that in some plans, almost 2% of the population has cancer and would be included in the measure if cancer was not excluded. This is a significant proportion of the population that could potentially impact the measure rates. No inferences about the hospice exclusion could be drawn because the majority of the plans could not identify exclusions.

2b4. 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 2b5.

2b4.1. What method of controlling for differences in case mix is used?
☐ **No risk adjustment or stratification**
☐ **Statistical risk model with** Click here to enter number of factors **risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use 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.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

2b4.4a. What were the statistical results of the analyses used to select risk factors?

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

2b4.5. 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 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

2b4.10. 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*)

2b4.11. 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*)

_____

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

To assess significant differences in measure rates, the data described in sections 1.5 and 1.6 above were used to calculate the mean, median, standard deviation, and interquartile range for the measure rates for the Medicare population and the Medicaid population. In addition, for the Medicaid population, the rates were divided into quartiles, and a Student's t-test was used to compare the rates between the bottom quartile (75th percentile) and top quartile (25th percentile). A student's t-test was used to compare the rates of the plans in the 25th percentile to the plans with rates in the 75th percentile. The statistics are for the Medicare population is reported below in 2b5.2, Tables 1 and 2. The statistics for the Medicaid population is reported below in 2b5.2, Tables 3 and 4.

2b5.2. 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 1. Variation in Measure Rates - Medicare Population (reported as number per 1,000 members)

Mean	Median	Standard Deviation
3.03	2.89	1.02

Table 2. Interquartile Range of Measure Rates - Medicare Population (reported as number per 1,000 members)

Minimum	1.94
25th Percentile	2.59
50th Percentile	2.89
75th Percentile	3.32
Maximum	4.41
Interquartile Range	0.73

Table 3. Variation in Measure Rates - Medicaid Population (reported as number per 1,000 members)

Mean	Median	Standard Deviation
2.68	2.38	1.80

Table 4. Interquartile Range of Measure Rates - Medicaid Population (reported as number per 1,000 members)

Minimum	0.70
25th Percentile	1.35
50th Percentile	2.38
75th Percentile	3.61
Maximum	6.12
Interquartile Range	2.26
Student's t-test p-value	0.085

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

For the Medicare population, the measure rates showed variation, with a standard deviation of 1.02 and an Interquartile Range was 0.73.

For the Medicaid population, the measure rates showed a greater variation, with a standard deviation of 1.80 and an Interquartile Range of 2.26.

There is a significant difference in measure rates between the top and bottom quartile of the plans included in the testing using the less conservative alpha of 0.10 because of the small measure rates (*P*=0.085 at alpha=0.10). This variation shows that there are meaningful differences in rates across plans.

2b6. 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 SDS 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 SDS 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.***

Only one set of specifications is provided for this measure.

2b6.1. 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*)

2b6.2. 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*)

2b6.3. 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*)

_____

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

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

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

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

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

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

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing*

data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data)

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

3. Feasibility

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

3a. Byproduct of Care Processes

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

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), Other
If other: Prescription claims data

3b. Electronic Sources

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

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*
ALL data elements are in defined fields in electronic claims

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

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.**
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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
Pilot test sites indicated the measure was feasible and results were able to be reported efficiently and accurately. CMS calculates the measure for Part D plans. The data is readily available (prescription claims data and medical data).

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
PQA develops and maintains numerous performance measures related to the medication use system. The Measures are the proprietary property of PQA, and it is in the interest of PQA to protect and promote the appropriate use of the Measures. PQA may approve an organization´s use of the Measures; however, no organization may use the Measures without first obtaining permission from PQA prior to using the Measures. Certain uses of the Measures are only approved with a licensing agreement from PQA that specifies the terms of use an the licensing fee. PQA reserves the right to determine the conditions under which it will approve and/or

license the Measures.

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. Current and Planned Use

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

Planned	Current Use (for current use provide URL)
Public Reporting Quality Improvement (Internal to the specific organization)	Quality Improvement with Benchmarking (external benchmarking to multiple organizations) CMS Medicare Part D - Patient Safety Reports http://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovGenIn/index.html

4a.1. For each CURRENT use, checked above, provide:

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

Name of program and sponsor: CMS Medicare Part D Drug Benefit
Purpose: Monitor Opioid use by Part D beneficiaries
Geographic area: National, approximately 38 million beneficiaries in Part D plans.

4a.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 was developed in 2015.

4a.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
The measure is currently being used in the Medicare Part D Overutilization Monitoring System to monitor the utilization of opioids for members with the Medicare drug benefit.

CMS has announced plans to move this measure into the 2019 Part D Display Measures, using data from 2017.

Reporting of results is not yet available.

4b. 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.

4b.1. Progress on Improvement. (Not required for initial endorsement unless available.)
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

N/A

4b.2. If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.
N/A

4c. 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).

4c.1. Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.
No unintended negative consequences to individuals or populations were identified during testing. This measure, Use of Opioids from Multiple Providers and at High Dosage in Persons Without Cancer, has been implemented by CMS Part D as part of the Overutilization Monitoring System beginning January, 2016. To date, no negative consequences have been identified.

However, concerns have been raised that prescribing changes such as dose reduction (without offering or arranging evidence-based treatment for patients with opioid use disorder) might be associated with unintended negative consequences, such as patients seeking heroin or other illicitly obtained opioids (1,2) or interference with appropriate pain treatment.(3) Data indicate that if access to prescription opioids is limited, some users of opioid analgesics will transition to heroin or other illicitly obtained opioids, leading to increased overdose death coincident with prescribing restrictions.(1) There are also concerns about pain being underdiagnosed and undertreated, particularly for ethnic and racial minorities, females, children or infirm elderly, or individuals who may be perceived to have mental health problems or are drug seeking.(4,5)

These concerns must be balanced by the current situation in the United States which has been described by the CDC as an epidemic of opioid abuse, overdose, and deaths. Overdose involving opioid analgesics killed almost 17,000 persons in 2010 and the number of people with opioid analgesic use disorders increased to nearly 2 million.(6) We believe the potential benefits of monitoring those patients receiving the very highest doses of opioids for extended periods of time or receiving these drugs from multiple providers outweighs potential negative consequences.

References:
1. Coffin P, Banta-Green C. The dueling obligations of opioid stewardship. Ann Intern Med. 2014;160:207–8. doi.org/10.7326/ M13-2781. (PMID 25133372).
2. Cicero,T, Ellis M, Harney J. Shifting Patterns of Prescription Opioid and Heroin Abuse in the United States. N Engl J Med. 2015; 373:1789-90. DOI: 10.1056/NEJMc1505541. (PMID 26510045).
3. Twillman RK, Kirch R, Gilson A. Efforts to control prescription drug abuse: Why clinicians should be concerned and take action as essential advocates for rational policy. CA Cancer J Clin. 2014;64:369–76. doi.org/10.3322/caac.21243. (PMID 25044063).
4. Kirschner N, Ginsburg J, Snyder LS, Health and Public Policy Committee of the American College of Physicians. Prescription Drug Abuse; executive summary of a policy position paper from the American College of Physicians. Ann Intern Med. 2014;160:198-200. doi:10.7326/M13-2209. (PMID 24323199).
5. Institute of Medicine (US) Committee on Advancing Pain Research, Care, and Education. Relieving Pain in America: A Blueprint for Transforming Prevention,Care, Education, and Research. Washington (DC): National Academies Press (US); 2011. (PMID: 22553896). Available at: http://www.nap.edu/read/13172/chapter/1
6. Coffin P, Banta-Green C. The dueling obligations of opioid stewardship. Ann Intern Med. 2014;160:207. doi: 10.7326/M13-2781. (PMID 24322334).

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

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 completely harmonized?

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.)
N/A

Appendix

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

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** PQA
Co.2 **Point of Contact:** Julie, Kuhle, jkuhle@pqaalliance.org, 515-554-6685-
Co.3 **Measure Developer if different from Measure Steward:** PQA
Co.4 **Point of Contact:** Julie, Kuhle, jkuhle@pqaalliance.org, 515-554-6685-

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

PQA Quality Metrics Expert Panel 2015
Responsible for review and consideration of the measure concept and all testing results of the draft measure
Bimal Patel*, Pharm D, MS MedImpact

Catherine Coast, PharmD Highmark
Chris DuPaul, MBA CVS Caremark
Christopher Dezii,RN, MBA, CPHQ Bristol-Myers Squibb
Christopher Powers, PharmD CMS
David Nau, PhD, RPh, CPHQ PQS
Gary Erwin, PharmD Omnicare
Gary Young, JD, PhD Northeastern University
Jenny Weber*, PharmD, MS, PCPS,CGP, BCACP Humana
Jessica Frank, PharmD OutcomesMTM
Karen Farris, PhD University of Michigan
Keith Widmer, RPh, BCPP Express Scripts
Kent Summers, RPh, PhD Astellas
Lynn Deguzman, PharmD, CGP Kaiser Permanente
Mary Ann Kliethermes, PharmD Midwestern University
Mitzi Wasik, PharmD, PCPS Coventry Health Care/Aetna
Pat Gleason, Pharm D, BCPS Prime Therapeutics
Steve Riddle, PharmD, BCPS Wolters Kluwer Health
Steven Burch, RPh, PhD GlaxoSmithKline
Tony Willoughby, PharmD HealthMart-McKesson
* denotes co-chair

PQA Medication Use Safety Workgroup 2014
Responsible for development of the measure
Amber Baybayan OutcomesMTM
David Belew MedHere Today
Rachael Boggs PQA Invited Guest Participant
Stay Bontha PerformRx
Sara Burnheimer UPMC Health Plan
Patrick Campbell University of Arizona College of Pharmacy
Scott Campbell PQA Invited Guest Participant
Rebecca Chater Ateb
Trina Clark GlaxoSmithKline
Victor Cohen American Society of Health-System Pharmacists (ASHP)
Michael Contos Indian Health Services
Karen Davidson Therapeutic Research Center (home of Pharmacist's Letter and Prescriber's Letter)
Shelly Delaville American Society of Consultant Pharmacists (ASCP)
James DeVita CVS/Caremark
Sara Ericsson MedImpact Healthcare Systems, Inc.
Marybeth Farquhar URAC
Alison Farrell Ahold USA
Cindi Fitzpatrick U.S. Food & Drug Administration (FDA)
Jeremy Fredell Express Scripts, Inc.
George Garmer CARE Pharmacies Cooperative
Jennifer Gatsos-Walter Wolters Kluwer Health, Clinical Solutions
Mary Ghods U.S. Food & Drug Administration (FDA)
James Glass Rite Aid
Averill Gordon Walgreen Co.
Lindsey Gumbo Pharmaceutical Research & Manufacturers of America (PhRMA)
Tracy Harrell SinfoníaRx
Tiffany Harris SCAN Health Plan
Shannon Harrison Highmark Health Services
Lisa Hines* University of Arizona College of Pharmacy
John Kessler National Alliance of State Pharmacy Associations (NASPA)
Mi´a Kirkland Wellcare
Nicholas Kostek Kaiser Permanente
Maribeth Kowalski Purdue Pharma, L.P.
Jason Kinsman RxAnte

Edward Lennard U.S. Office of Personnel Management
Patricia Marchlowska Lilly USA
Peter Marshall HealthPartners
Kevin Masci Target
Richard McLeod Pfizer, Inc.
Diane McNally Centers for Medicare & Medicaid Services (CMS)
Brent Merrick Cigna-HealthSpring
Leslie Miller Gorman Health Group
Joel Montavon Catamaran
Kim Moon PQA Invited Guest Participant
Gina Moore PQA Invited Guest Participant
Scott Nakagawa Applied Research Works
Patricia Neafsey ActualMeds Corporation
Jeffrey Nesheim Takeda Pharmaceuticals America, Inc.
Michael Nguyen CenseoHealth
Kyle Null* University of Mississippi Center for Pharmaceutical Marketing & Management
Udo Nwachukwu Mirixa Corporation
Steven Oh Health Mart Systems Inc.
Maria Osborne American Pharmacists Association (APhA)
Nicole Paterson Fairview Medication Therapy Management
Jacqui Pesa Johnson & Johnson
Roger Pinsonneault RelayHealth
Richard Segal University of Florida College of Pharmacy
Bupendra Shah Long Island University Arnold & Marie Schwartz College of Pharmacy
Christine Sommer First DataBank
Catherine Starner Prime Therapeutics
Karen Stockl UnitedHealth Group
Brian Sweet AstraZeneca
Christie Teigland Inovalon, Inc.
Jennifer Thomas National Alliance of State Pharmacy Associations (NASPA)
Ly Tran PharmMD
Maria Vassilakis Astellas Scientific and Medical Affairs, Inc.
Kathleen Vest PQA Invited Guest Participant
Brandi Rosberg Walmart
Jennifer Weber Humana
Elizabeth Whaley-Buono MeadWestvaco
Jennifer Williams Aetna
Melissa Wilson Capital Health Plan

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2015

Ad.3 **Month and Year of most recent revision:** 05, 2015

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

Ad.5 **When is the next scheduled review/update for this measure?** 10, 2016

Ad.6 **Copyright statement:**

Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 2983
De.2. **Measure Title:** Potassium Sample Hemolysis in the Emergency Department
Co.1.1. **Measure Steward:** Cleveland Clinic
De.3. **Brief Description of Measure:** Percentage of laboratory potassium samples drawn in the emergency department (ED) with hemolysis.
1b.1. **Developer Rationale:** Hemolysis is the rupture red blood cells with a release of hemoglobin and other intracellular content into plasma interfering with multiple laboratory tests including potassium. Hemolyzed samples account for the majority of rejected samples. The American Society of Clinical Pathology consider a hemolysis rate below 2% best practice (Lowe G, Stike R, Pollack M, Bosley J, O´Brien P, Hake A, et al. Nursing blood specimen collection techniques and hemolysis rates in an emergency department: analysis of venipuncture versus intravenous catheter collection techniques. J Emerg Nurs 2008;34:26-32.) The Emergency Department accounts for a large proportion of a hospital's labs rejected specimens for hemolysis.

Heyer, N. J., Derzon, J. H., Winges, L., Shaw, C., Mass, D., Snyder, S. R., et al. (2012). Effectiveness of practices to reduce blood sample hemolysis in EDs: A laboratory medicine best practices systematic review and meta-analysis. Clinical Biochemistry, 45(13–14), 1012-1032.

S.4. **Numerator Statement:** ED Potassium Samples with Hemolysis
S.7. **Denominator Statement:** all ED patients getting a lab potassium sample
S.10. **Denominator Exclusions:** None

De.1. **Measure Type:** Intermediate Clinical Outcome
S.23. **Data Source:** Electronic Clinical Data : Laboratory
S.26. **Level of Analysis:** Facility

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

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

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

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

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *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.

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 presents a number of studies that demonstrate that hemolysis is preventable by using appropriate blood draw techniques. The evidence is weak to moderate and several studies provided are rated as insufficient evidence.

Guidance from the Evidence Algorithm

1-No→3-Yes→4-Yes→5a or 5b

Questions for the Committee:

If the developer provided updated evidence for this measure:

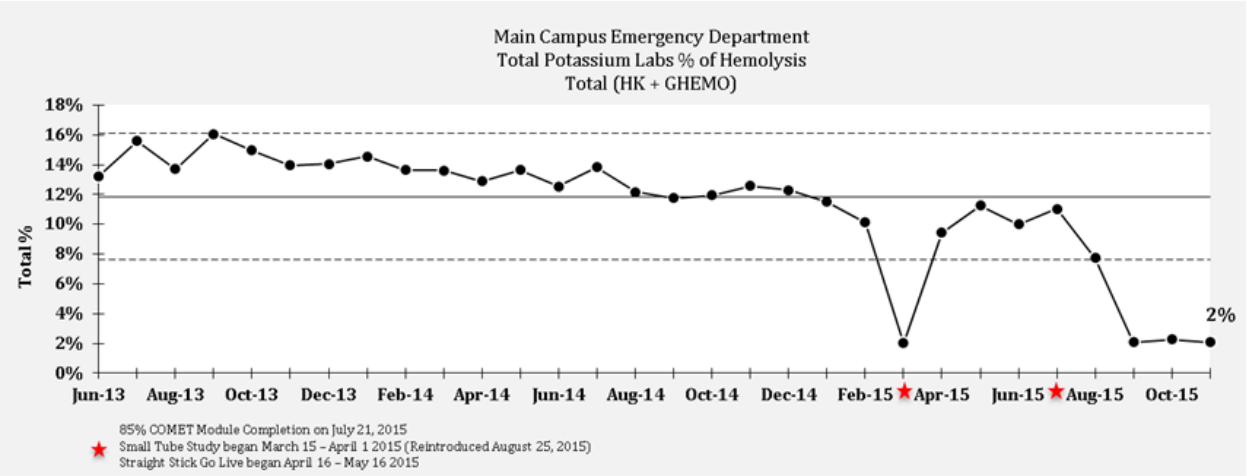
- *For process measures:*
  - *What is the relationship of this measure to patient outcomes?*
  - *How strong is the evidence for this relationship?*
  - *Is the evidence directly applicable to the process of care being measured?*
  -

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

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

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

- The developer includes a graph that depicts emergency room hemolysis rates from potassium lab samples in June-2013 to October 2015 at the Cleveland Clinic. The percentage of hemolysis seem to have decreased over time with about 13% hemolysis rate in June-2013 and a 2% rate in October 2015. The rates appear to have steadily reduced overtime.



Disparities

- The developer states that there are no published studies that have identified disparities in ED hemolysis. The problems appears to be consistent across populations.

Questions for the Committee:

- *Specific question on information provided for gap in care.*
- *Is there a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*

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)

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s):

Specifications:

Questions for the Committee :

- *Specific questions on the specifications, codes, definitions, etc.*
- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

eMeasure Technical Advisor(s) review (if not an eMeasure, delete this section):

Submitted measure is an HQMF compliant eMeasure	The submitted eMeasure specifications follow the industry accepted format for eMeasure (HL7 Health Quality Measures Format (HQMF)). **HQMF specifications** ☒ **Yes** ☐ **No**
Documentation of HQMF or QDM limitations	N/A – All components in the measure logic of the submitted eMeasure are represented using the HQMF and QDM; **Measure logic tested with Bonnie and in Epic 2014**
Value Sets	The submitted eMeasure specifications uses existing value sets when possible and uses new value sets that have been vetted through the VSAC
Measure logic is unambiguous	Submission includes test results from a simulated data set demonstrating the measure logic can be interpreted precisely and unambiguously, as well as results from a single ONC certified EHR
Feasibility Testing	The feasibility analysis submitted by the measure developer meets the requirements to be

	considered for eMeasure Trial Approval.

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☐ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☐ **Yes** ☐ **No**

Method(s) of reliability testing [method of reliability testing]

Results of reliability testing [Results of reliability testing]

Guidance from the Reliability Algorithm [Algorithm guidance]

Questions for the Committee:

- *Specific questions on the method and results of reliability testing.*
- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

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

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☐ **Yes** ☐ **Somewhat** ☐ **No**
Specification not completely consistent with evidence

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

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, summarize the validity testing from the prior review:

SUMMARY OF TESTING

Validity testing level ☐ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

☐ **Face validity only**

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

Validity testing method:

Validity testing results:

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*
- *Other specific question of the validity testing?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

[Summarize and analysis of exclusions]

Questions for the Committee:

- *Specific questions on the information about exclusions.*
- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☐ **None** ☐ **Statistical model** ☐ **Stratification**

Conceptual rationale for SDS factors included ? ☐ **Yes** ☐ **No**

SDS factors included in risk model? ☐ **Yes** ☐ **No**

Risk adjustment summary [Risk adjustment summary

Questions for the Committee:

- *Specific questions on the risk-adjustment approach.*
- *If a justification for no risk adjustment is provided, is there any evidence that contradicts the developer's rationale and analysis?*
- *Is an appropriate risk-adjustment strategy included in the measure?*
- *Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented?*
- *Are all of the risk adjustment variables present at the start of care? If not, describe the rationale provided.*
- *Do you agree with the developer's rationale that there is no conceptual basis for adjusting this measure for SDS factors?*
- *Do you agree with the developer's decision, based on their analysis, to not include SDS factors in their risk-adjustment model?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

[meaningful differences]

Question for the Committee:

○ *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

[comparability of data sources]

2b7. Missing Data

[missing data]

Committee pre-evaluation comments

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

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.

- This measure is generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score)
- ALL data elements are in defined fields in electronic health records (EHRs).
- Feasibility Assessment Scorecard Included.
- There are multiple ways to collect this data. The developer collected data from both the ONC certified EMR Epic(Epic 14) and the ONC certified Laboratory information systems(LIS) (Sunquest 7.2).
- Obtaining data from the EMR was challenging, perhaps because the lab add on(LIS) from Epic was not purchased.
- Data required a few iterations before sufficient data was obtained.
- The developer chose to go with our current LIS for lab information Sunquest. Since this information can be obtained directly from the LIS we presumed this to be their reference value for ED hemolysis.
- Preliminary data analysis from our EMR vs LIS(both Onc certified) for about 70,000 patients a year for 2 years with about 35,000 lab potassium results for ED patients showed hemolysis rates that were close but not an exact match likely do to data definitions and population definitions.
- The developer has plans to submit request for funding/grant to analyze the variance which is presumed to be more related to their data ask.
- No fees or licensing requirements to use any aspect of the measure as specified, were reported.

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

Criterion 4: Usability and Use

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

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

Current uses of the measure:]
Publicly reported? ☐ **Yes** ☒ **No**

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

Accountability program details:

- This measure is not currently in use.
- ED hemolysis has significnat impact on care of ED patients. Since most of the issue around this causes are preanaltycia; Its impact on ED patients and work in both ED and lab lead significantly increased work around redraw and re-testing.
- Panned use includes: Public Reporting, Public Health/Disease Surveillance, Quality Improvement with Benchmarking (external benchmarking to multiple organizations), and Quality Improvement (Internal to the specific organization)

Improvement results:

- Organizations that have formally addressed the ED hemolysis problem (including Cleveland Clinic publications pending) have seen improvement in hemolysis rates by addressing some of the pre analytical issues some by either going to straight stick for most/all of ED blood samples or modification of equipment(use small volume/vacuum).
- Low vacuum and discard tubes reduce hemolysis in samples drawn from intravenous catheters.
- Reducing blood sample hemolysis at a tertiary hospital emergency department.

Potential harms:

- Developer did not identify any unintended consequences related to this measure.

Feedback :

- Developer did not identify any specific feedback loops related to this 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

Criterion 5: Related and Competing Measures

Related or competing measures
N/A

Harmonization
N/A

Pre-meeting public and member comments

-

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: **Potassium Sample Hemolysis in the Emergency Department**
IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: 5/6/2016

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

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

Outcome

☐ Health outcome: Click here to name the health outcome

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☒ Intermediate clinical outcome (*e.g., lab value*): Potassium Sample Hemolysis in the Emergency Department

☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

_____

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

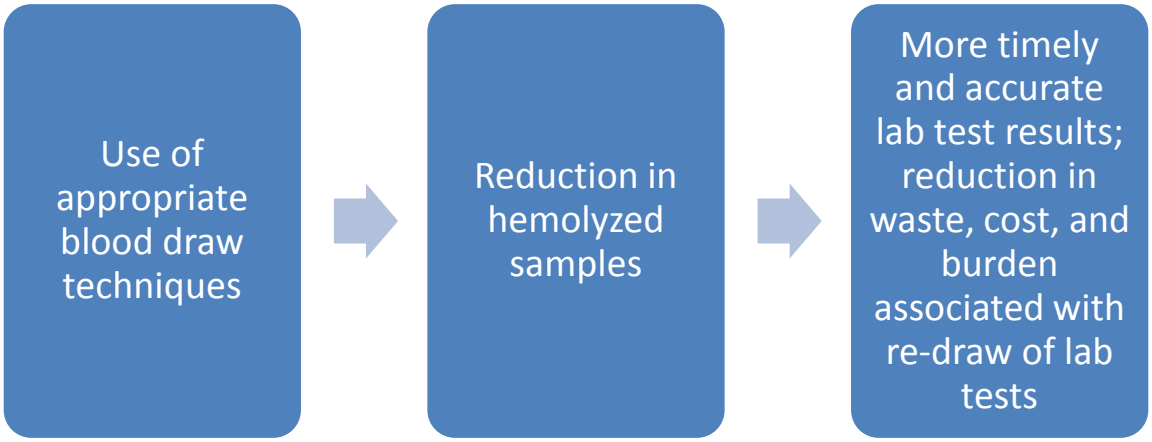
1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

_____

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.



1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?
☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***
☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***
☒ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***
☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION
1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

Proehl JA, Bradford JY, Leviner S, et al. Emergency Nursing Resources Development Committee. Clinical practice guideline: prevention of blood specimen hemolysis in peripherally-collected venous specimens. 2012. Available at https://www.ena.org/practice-research/research/CPG/Documents/HemolysisCPG.pdf

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

Guideline recommendations 1-18 on page 8.

"1. Education of the staff performing phlebotomy may decrease hemolysis. **Level C – Weak** (Halm, 2009; Ong, 2009; Corkill, 2012)
2. The type of personnel performing phlebotomy does not influence hemolysis. **Level C – Weak** (Halm, 2009; Bush, 2010; Harrison, 2010; Saleem, 2009; Ong, 2008)
3. Hemolysis is less likely when blood is drawn from the antecubital fossa. **Level B – Moderate** (Tanabe, 2003; Fang, 2008; Heyer, 2012)
4. Minimize tourniquet time by removing the tourniquet after identifying the venipuncture site while preparing equipment and as soon as good blood flow is established. **Level C – Weak** (Saleem, 2009)
5. There is insufficient evidence to determine if the number of venipuncture attempts affects hemolysis. **Level – I/E** (Saleem, 2009)

6. There is insufficient evidence as to whether intravenous catheter insertion perceived to be difficult is associated with an increased risk of hemolysis. **Level – I/E** (Stauss, 2012; Ong, 2008)
7. Direct venipuncture with straight needles is less likely to cause hemolysis than blood collection through intravenous catheters. **Level B – Moderate** (Tanabe, 2003; Ong, 2009; Bush, 2010; Berger-Achituv, 2010; Heyer, 2012; Saleem, 2009)
8. Stainless steel needles are less likely to cause hemolysis than intravenous catheters; teflon catheters are less likely to cause hemolysis than Vialon™ catheters. **Level C – Weak** (Raisky, 1994; Sharp, 1998)
9. There is conflicting evidence regarding the influence of needle or catheter gauge on hemolysis. Level: I/E. (Sharp, 1998; Sharp, 2003; Tanabe, 2003; Sequin, 2004; Heyer, 2012)
10. There is conflicting evidence regarding hemolysis with syringes versus vacuum tubes. **Level – I/E** (Sharp, 2003; Halm, 2009; Ong, 2009; Bush, 2010; Saleem, 2009; Sequin, 2004)
11. Drawing blood through an extension tubing attached to an intravenous catheter does not increase hemolysis in adults. **Level C – Weak** (Stauss, 2012)
12. Drawing blood through needleless connectors does not increase hemolysis. **Level – Moderate** (Dwyer, 2006; Sharp, 2003)
13. There is insufficient evidence regarding the impact of the rate of blood flow into a vacuum tube on hemolysis. **Level – I/E** (Ong, 2008)
14. Low (partial) vacuum tubes result in less hemolysis. **Level B – Moderate** (Heyer, 2012; Schwartzer, 2001)
15. Filling vacuum tubes to their recommended volume decreases hemolysis. **Level C – Weak** (Unger, 2007; Tamechika, 2006)
16. Properly functioning pneumatic tube systems do not increase hemolysis. **Level C – Weak** (Stair, 1995; Fang, 2008; Ellis, 2009; Saleem, 2009; Streichert, 2011; Evilyaoglu, 2012)
17. There is insufficient evidence to determine if the volume of venipunctures performed influences hemolysis. **Level – I/E** (Hawkins, 2010)
18. There is insufficient evidence to determine if monitoring hemolysis rates and providing feedback to the staff performing phlebotomy decreases the incidence of hemolysis. **Level – I/E** (McGrath, 2012)"

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

See 1a.4.2 for individual recommendation grades. Summary of definitions of grades below:

Table 1. Levels of Recommendation for Practice

Level A recommendations: High
- Reflects a high degree of clinical certainty
- Based on availability of high quality level I, II and/or III evidence available using Melnyk & Fineout-Overholt grading system (Melnyk & Fineout-Overholt, 2005)
- Based on consistent and good quality evidence; has relevance and applicability to emergency nursing practice
- Is beneficial

Level B recommendations: Moderate
- Reflects moderate clinical certainty
- Based on availability of Level III and/or Level IV and V evidence using Melnyk & Fineout-Overholt grading system (Melnyk & Fineout-Overholt, 2005)
- There are some minor or inconsistencies in quality evidence; has relevance and applicability to emergency nursing practice
- Is likely to be beneficial

Level C recommendations: Weak
- Level V, VI and/or VII evidence available using Melnyk & Fineout-Overholt grading system (Melnyk & Fineout-Overholt, 2005) - Based on consensus, usual practice, evidence, case series for studies of treatment or screening,

anecdotal evidence and/or opinion
 There is limited or low quality patient-oriented evidence; has relevance and applicability to emergency nursing practice
 Has limited or unknown effectiveness

Not recommended for practice
 No objective evidence or only anecdotal evidence available; or the supportive evidence is from poorly controlled or uncontrolled studies
 Other indications for not recommending evidence for practice may include:
- Conflicting evidence
- Harmfulness has been demonstrated
- Cost or burden necessary for intervention exceeds anticipated benefit
- Does not have relevance or applicability to emergency nursing practice

 There are certain circumstances in which the recommendations stemming from a body of evidence should not be rated as highly as the individual studies on which they are based. For example:
- Heterogeneity of results
- Uncertainty about effect magnitude and consequences,
- Strength of prior beliefs
- Publication bias

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

See question 1a4.3 for definitions of all grades.

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

Christenson RH, Snyder SR, Shaw CS, Derzon JH, Black RS, Mass D, et al. Laboratory medicine best practices: systematic evidence review and evaluation methods for quality improvement. Clin Chem 2011;57:816-25.

Melnyk, B.M. & Fineout-Overholt, E. (2005). Evidence-Based Practice in Nursing & Healthcare. A Guide to Best Practice. Lippincott, Williams & Wilkins

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☒ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

Heyer NJ, Derzon JH, Winges L, et al. Effectiveness of practices to reduce blood sample hemolysis in EDs: a laboratory medicine best practices systematic review and meta-analysis. *Clin Biochem.* 2012;45:1012-1032

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Christenson RH, Snyder SR, Shaw CS, Derzon JH, Black RS, Mass D, et al. Laboratory medicine best practices: systematic evidence review and evaluation methods for quality improvement. Clin Chem 2011;57:816-25.

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

ED lab sample hemolysis

ENA Guideline:

Pre-analytic variables related to peripheral venous specimen collection and transportation decrease blood culture hemolysis

Heyer Review: Emergency department (ED) practices for reducing hemolysis in blood samples sent to the clinical laboratory for testing.

.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

ENA Guideline: The guideline graded the quality of evidence for each recommendation as follows See question 1a.4.3 for definitions of grades:

1. Education of the staff performing phlebotomy may decrease hemolysis. **Level C – Weak** (Halm, 2009; Ong, 2009; Corkill, 2012)
2. The type of personnel performing phlebotomy does not influence hemolysis. **Level C – Weak** (Halm, 2009; Bush, 2010; Harrison, 2010; Saleem, 2009; Ong, 2008)
3. Hemolysis is less likely when blood is drawn from the antecubital fossa. **Level B – Moderate** (Tanabe, 2003; Fang, 2008; Heyer, 2012)

4. Minimize tourniquet time by removing the tourniquet after identifying the venipuncture site while preparing equipment and as soon as good blood flow is established. **Level C – Weak** (Saleem, 2009)
5. There is insufficient evidence to determine if the number of venipuncture attempts affects hemolysis. **Level – I/E** (Saleem, 2009)
6. There is insufficient evidence as to whether intravenous catheter insertion perceived to be difficult is associated with an increased risk of hemolysis. **Level – I/E** (Stauss, 2012; Ong, 2008)
7. Direct venipuncture with straight needles is less likely to cause hemolysis than blood collection through intravenous catheters. **Level B – Moderate** (Tanabe, 2003; Ong, 2009; Bush, 2010; Berger-Achituv, 2010; Heyer, 2012; Saleem, 2009)
8. Stainless steel needles are less likely to cause hemolysis than intravenous catheters; teflon catheters are less likely to cause hemolysis than Vialon™ catheters. **Level C – Weak** (Raisky, 1994; Sharp, 1998)
9. There is conflicting evidence regarding the influence of needle or catheter gauge on hemolysis. Level: I/E. (Sharp, 1998; Sharp, 2003; Tanabe, 2003; Sequin, 2004; Heyer, 2012)
10. There is conflicting evidence regarding hemolysis with syringes versus vacuum tubes. **Level – I/E** (Sharp, 2003; Halm, 2009; Ong, 2009; Bush, 2010; Saleem, 2009; Sequin, 2004)
11. Drawing blood through an extension tubing attached to an intravenous catheter does not increase hemolysis in adults. **Level C – Weak** (Stauss, 2012)
12. Drawing blood through needleless connectors does not increase hemolysis. **Level – Moderate** (Dwyer, 2006; Sharp, 2003)
13. There is insufficient evidence regarding the impact of the rate of blood flow into a vacuum tube on hemolysis. **Level – I/E** (Ong, 2008)
14. Low (partial) vacuum tubes result in less hemolysis. **Level B – Moderate** (Heyer, 2012; Schwartzer, 2001)
15. Filling vacuum tubes to their recommended volume decreases hemolysis. **Level C – Weak** (Unger, 2007; Tamechika, 2006)
16. Properly functioning pneumatic tube systems do not increase hemolysis. **Level C – Weak** (Stair, 1995; Fang, 2008; Ellis, 2009; Saleem, 2009; Streichert, 2011; Evilyaoglu, 2012)
17. There is insufficient evidence to determine if the volume of venipunctures performed influences hemolysis. **Level – I/E** (Hawkins, 2010)
18. There is insufficient evidence to determine if monitoring hemolysis rates and providing feedback to the staff performing phlebotomy decreases the incidence of hemolysis. **Level – I/E** (McGrath, 2012)

Heyer Review: The Heyer review graded the strength of evidence for individual practices as follows:

Straight needle venipuncture: High
Antecubital site vs distal site: High
Use of syringe vs vacuum tubes: Insufficient
Use of ≤21-gauge (larger) needles: Insufficient
Use of low (partial) vacuum tubes: Suggestive

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

ENA Guideline: See question 1a.4.3 for grade definitions

Heyer Review: The strength of evidence grades are defined as follows:

HIGH

Adequate volume of consistent evidence of substantial healthcare quality impact from studies without major limitations.

MODERATE

Some evidence of consistent substantial healthcare quality impact from studies without major limitations; OR an adequate volume of consistent evidence of moderate healthcare quality impact from studies without major limitations.

SUGGESTIVE

Limited evidence of moderate healthcare quality impact from a small number of studies without major limitations; OR the quality of some studies' design and/or conduct is limited.

INSUFFICIENT

Any estimate of an effect on healthcare quality impact is too uncertain.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: 1998-2012

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

ENA: The ENA guideline reviewed 9 randomized trials, 2 systematic review/meta-analyses, 6 prospective studies, 10 observational studies

Heyer:

The Heyer review included 11 cross-sectional/observational studies, 2 prospective trials, and 3 randomized trials.

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

ENA guideline: The ENA guideline does not discuss the certainty or confidence in the estimates of effect.

Heyer review: The Heyer review discusses the possibility that not all variation in hemolysis rates is necessarily attributable to the practice of interest in the reviewed studies due to insufficient control of external variables. This could possibly increase error variation in outcome estimates. However, they do not expect this variation to bias the overall estimate of the effectiveness of the practices of interest.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

ENA Guideline:

The ENA guideline does not provide an estimate of benefit across studies.

Heyer:

Straight needle venipuncture: the overall reduction in hemolysis from using straight needle venipuncture is consistently supported by the evidence, significant, and equal to about 84% (RR=0.16, 95% CI=0.11–0.24; see Fig. 3).

Antecubital site vs distal sites:

Based on these four studies, the

overall expected reduction is hemolysis of 55% (RR=0.45, 95% CI= 0.35–0.57) and the results are homogeneous ($Q_{Overall}$=2.20, p= 0.533, I^2=0.00) (Fig. 4).

Syringe vs vacuum tubes: The meta-analysis results for syringe effectiveness are heterogeneous ($Q_{Overall}$=19.29, p=0.00, I^2=89.63), with a reduction in hemolysis from use of a syringe of approximately 3% and not statistically signi!cantly different from no effect versus the comparison practice (RR=0.97, 95% CI= 0.81–1.17).

Use of ≤21-gauge (larger) needles: Although the meta-analysis mean risk ratio for !21 gauge (larger) needles is substantial (RR=0.37, 95% CI=0.27–0.52) and equal to approximately a 63% reduction in hemolysis, the individual study effect size results for needle size are "inconsistent" and heterogeneous ($Q_{Overall}$= 14.82, p=0.001, I^2=86.50)

Low (partial) vacuum tubes: The meta-analysis (Fig. 7) mean effect size rating for the two studies is equal to a reduction in hemolysis of approximately 89% (RR=0.11, 95% CI=0.02–0.52). Although the effect size results from the two studies were "consistent," they are heterogeneous (Q=4.66, p=0.03, I2=78.54).

There was not sufficient evidence to evaluate the following: longer vs shorter tourniquet time and phlebotomist vs ED medical staff.

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?
ENA: Harms associated with straight needle venipuncture included: more needle sticks for the patient and therefore more pain and anxiety, staff members have more potential for exposure to blood, and the laboratory specimens are generally obtained less quickly.

Heyer: No harms were evaluated as part of this review.

Neither study found evidence of serious harms associated with the use of the recommended techniques. As such we expect there to be a net benefit of using these techniques for blood draws.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

"Use of plasma collection tubes with smaller volumes and decreased vacuum significantly reduces hemolysis in ED blood samples" ADD CITATION Phelan MP, Reinecks EZ, et al.
Annals of Emergency Medicine Oct. 2015 abstract

This study found that small volume tubes significantly reduced hemolysis as compared to standard tubes. This supports the recommendation about small volume/vacuum tubes made in the reviews cited above, that had insufficient data to support use.

"One poke or two? Can intravenous catheters provide an acceptable blood sample? A dataset presentation, review of previous datasets and discussion" Dietrich H. Nov 2014 40(6). Journal of Emergency Nursing.

This paper found that use of IV to draw blood can also be done with low hemolysis rates. They did not know why they had low rates. In a letter to editor to be published from our group we found out that they were using low volume/vacuum tubes during the study period and this may have contributed to their low hemolysis rates. (please see letter to editor below, we can send a copy once it has been published)

Investigation supports the use of blood samples drawn from IV starts, regardless of site, without increasing risk of hemolysis.

We have written a letter to editor to of Journal of Emergency nursing responding to Dietrich's finding that should be published this summer:

Two articles came out last year about the variability and quality practices around this topic:

Howanitz PJ, Lehman CM, Jones BA, Meier FA, Horowitz GL. Practices for identifying and rejecting hemolyzed specimens are highly variable in clinical laboratories. *Arch Pathol Lab Med*. 2015;139(8):1014–1019.

Howanitz PJ, Lehman CM, Jones BA, Meier FA, Horowitz GL. Clinical laboratory quality practices when hemolysis occurs. *Arch Pathol Lab Med*. 2015;139(7):901–906.

We recently had a response to these accepted for publication as Letter to the editor for Arch Pathology and Lab Medicine describing the cost associated with hemolysis. If published soon will provide a copy.

We have published abstracts and are in the process of writing 2 papers on the topic that might be available when the committee meets we will keep you posted

Phelan MP, Reineks EZ, Schold, J, Podolsky, SR, Schmidt, J, Hustey FM, Meldon S, Barbour T, Regotti K, ProcopGW. Does Pneumatic Tube System Transport Contribute to Hemolysis Rates in ED Blood Samples? SAEM AnnualMeeting, San Diego, California, May 2015. Acad Emerg Med, 22 Apr 2015. doi: 10.1111/acem.12656.

Support the contention that this is a pre analytical problem related to the ED method of drawing not transportation issue or lab problem.

Phelan MP, Reineks EZ, Berriochoa JP, Hustey FM, Podolsky SR, Meldon SW, Kovach A, Schmidt JA, Schold JD, Mcclintock PE, Procop GW. Use of plasma collection tubes with smaller volumes and decreased vacuum significantly reduces hemolysis in ED blood samples. Cleveland Clinic and MetroHealth Medical Center, Cleveland, Ohio, October 2015. doi: 10.10.16/j.annemergmed.2016.07.098.
Studied the impact of switching to small voullum/vaccumm tubes on ED lab sample hemolysis(reduced it)
Support use of a small volume/vacuum tubes when drawing blood from an IV to reduce hemolysis

Phelan MP, et al. Impact of Hemolyzed Blood Specimens on Emergency Department Patient Throughput. SAEM 2016 Annual Meeting, New Orleans, LA, 2016.
Studied one quarter of ED lab data and assessed the impact on throughput from hemolyzed samples(increased throughput significantly)

Phelan MP, et al. Validation and Evaluation of Pre-analytical Factors Associated with Hemolysis in ED Blood Samples. SAEM 2016 Annual Meeting, New Orleans, LA, 2016.

Evaluated 1 years' worth of data assessed the same pre analytical factors in Heyer. (Validated)
This study confirms previous findings that the use of straight needles and the antecubital location is significantly associated with reduced hemolysis, supporting the original conclusion of Heyer and colleagues.[4] In addition, our findings indicate that shorter tourniquet time (less than 60 seconds) and the use of larger-gauge needles for IV draws were significantly associated with lower hemolysis. No association was found between syringe versus vacuum tube sample collection in regards to incidence of hemolysis. These findings confirm and support best practices in the ED to reduce hemolysis and improve efficiencies in the acquisition of blood samples.

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

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

Graph_depicting_ED_hemolysis_over_time_during_performance_improvement_project.docx,Hemolysis_NQF_Evidence_Attachment_5_06_fin.docx

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., the benefits or improvements in quality envisioned by use of this measure)*
Hemolysis is the rupture red blood cells with a release of hemoglobin and other intracellular content into plasma interfering with multiple laboratory tests including potassium. Hemolyzed samples account for the majority of rejected samples. The American Society of Clinical Pathology consider a hemolysis rate below 2% best practice (Lowe G, Stike R, Pollack M, Bosley J, O´Brien P, Hake A, et al. Nursing blood specimen collection techniques and hemolysis rates in an emergency department: analysis of venipuncture versus intravenous catheter collection techniques. J Emerg Nurs 2008;34:26-32.) The Emergency Department accounts for a large proportion of a hospital's labs rejected specimens for hemolysis.

Heyer, N. J., Derzon, J. H., Winges, L., Shaw, C., Mass, D., Snyder, S. R., et al. (2012). Effectiveness of practices to reduce blood sample hemolysis in EDs: A laboratory medicine best practices systematic review and meta-analysis. Clinical Biochemistry, 45(13–14), 1012-1032.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
see attachement graph depicting hemolysis over time files under 1a

We wer bale to pull data from our Lab Information System, Sunquest and provide monthy data on our hemolyis incidence.

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.**
Hemolyzed blood samples are frequently received in clinical laboratories, comprising as much as 3.3% of all
Routine samples and accounting for up to 40%–70% of all unsuitable samples identified — nearly five times higher than other causes, such as insufficient, incorrect, and clotted samples [1]. The American Society for Clinical Pathology established a 2% or lower benchmark for hemolysis rates among laboratory blood samples [2]. Hospital EDs
have been identified as a major source of hemolyzed samples. Two studies in hospital EDs found hemolysis rates of more than 30% [3,4], while many others observed rates (ranging from 6.8 to 19.8%) that were considerably higher than the established benchmark

[5–9]. Several studies [4,8,9] identified ED hemolysis rates that were significantly elevated compared to other hospital departments.

[1] Lippi G, Blanckaert N, Bonini P, Green S, Kitchen S, Palicka V, et al. Haemolysis: an overview of the leading cause of unsuitable specimens in clinical laboratories. Clin Chem Lab Med 2008;46:764-72.

[2] Lowe G, Stike R, Pollack M, Bosley J, O´Brien P, Hake A, et al. Nursing blood specimen collection techniques and hemolysis rates in an emergency department: analysis of venipuncture versus intravenous catheter collection techniques. J Emerg Nurs 2008;34:26-32.

[3] Grant M. The effect of blood drawing techniques and equipment on the hemolysis of ED laboratory blood samples. J Emerg Nurs 2003;29:116-21.
[4] Soderberg J, Jonsson PA, Wallin O, Grankvist K, Hultdin J. Haemolysis index—an estimate of preanalytical quality in primary health care. Clin Chem Lab Med 2009;47:940-4.

[5] Burns ER, Yoshikawa N. Hemolysis in serum samples drawn by emergency department personnel versus laboratory phlebotomists. Lab Med 2002;33:378-80.
[6] Dwyer DG, Fry M, Somerville A, Holdgate A. Randomized, single blinded control trial comparing haemolysis rate between two cannula aspiration techniques. Emerg Med Australas 2006;18:484-8.
[7] Ong ME, Chan YH, Lim CS. Observational study to determine factors associated with blood sample haemolysis in the emergency department. Ann Acad Med Singapore 2008;37:745-8.
[8] Pretlow L, Gandy T, Leibach EK, Russell B, Kraj B. A quality improvement cycle: hemolyzed specimens in the emergency department. Clin Lab Sci 2008;21:219-24.
[9] Tanabe P, Kyriacou DN, Garland F. Factors affecting the risk of blood bank specimen hemolysis. Acad Emerg Med 2003;10:897-900.

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*
There are no published studies, its a uniformly distrubuted problem across all populations.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
There are no published studies, its a uniformly distrubuted problem across all populations.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, Frequently performed procedure, High resource use
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
CDC and Emergency Nurse Association have publsihed material on the importnace of this topic that have been cited.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
Heyer, N. J., Derzon, J. H., Winges, L., Shaw, C., Mass, D., Snyder, S. R., et al. (2012). Effectiveness of practices to reduce blood sample hemolysis in EDs: A laboratory medicine best practices systematic review and meta-analysis. Clinical Biochemistry, 45(13–14), 1012-1032.

https://www.ena.org/practice-research/research/CPG/Documents/HemolysisCPG.pdf

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide**

evidence that the target population values the measured PRO and finds it meaningful. *(Describe how and from whom their input was obtained.)*
not applicable

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria 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. **Cross Cutting Areas** *(check all the areas that apply):*
Safety

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.)*
The measure specifications are included as an attachment with this submission. Value set details at VSAC webpage: https://vsac.nlm.nih.gov/

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 an eMeasure **Attachment:** HEMOLYSISinED_v4_Artifacts_08282015.zip

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** Potassium_Sample_Hemolysis_in_the_Emergency_Departmentfin2_-6-_highlights.pdf

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
not applicable

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
ED Potassium Samples with Hemolysis

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Each sample drawn during 12 consecutive months

S.6. **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, 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.
patients with lab potassium sample where the result was hemolyzed.

Please see attached specifications

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*

all ED patients getting a lab potassium sample

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Children's Health, Maternal Health, Populations at Risk, Populations at Risk : Dual eligible beneficiaries, Populations at Risk : Individuals with multiple chronic conditions, Populations at Risk : Veterans, Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
All ED patient who get lab potassium sample

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
None

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
not applicable

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
Not applicable.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Not applicable.

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Provided in response box S.15a

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
Not applicable.

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **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.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
The total number of hemolized potassiun samples are divided by the total number of ED potassium samples

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Not applicable.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not applicable.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Missing data for this measure is expected to be minimal and should not impact measure performance.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data : Laboratory

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Not applicable

S.25. **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.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility, Other
If other: emergency department

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
Not applicable.

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
CMSv0_bonnie_testing_April_19_2016.xlsx,Bonnie_measure_testing.docx

National Quality Forum

Measure Testing Form for Trial Approval Program

Measure Title: **Emergency Department blood sample Potassium hemolysis**

Date of Submission: April 15, 2015

Type of Measure:

☐ Composite

☐ Outcome (*including PRO-PM*)

☐ Cost/resource

x Process

x Efficiency

☐ Structure

Instructions

- A measure submision that is to be considered for the Trial Approval Program must complete this form in its entirety. Either a test data set provided by the measure developer, or the use of the Bonnie tool is acceptable to provide prelminary testing results,
- **For <u>all</u> measures being submitted for potential acceptance into the Trial Approval Program, each section <u>must be filled out as completely as possible.</u>**
- Respond to <u>all</u> questions as instructed with answers immediately following the question. All information on testing of either a sample data set or results from Bonnie testing that can demonstrate, to the extent possible, the the measure meets the reliability and validity must be in this form..
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions at trialmeasures@qualityforum.org

<u>DATA and SAMPLING INFORMATION</u>

1. DATA/SAMPLE USED FOR <u>PRELMINARY</u> TESTING OF THIS MEASURE

It is important that the measure developer use a data set to conduct preliminary testing in order to evauate the measure logic and the inclusions/exclusions for the population used in the measure.

- **What type of data was used for testing**? (*The measure developer must provide a test data set that will provide some initial information to be used for the evaluation, or the Bonnie testing tool can use can be used to create a sample data set using synthesized patients.)* Please indicate whether the test data set used was provided through the measure developer, or through the Bonnie tool.

We tested the measure through Bonnie testing as well as at our own institution. Hemolysis rates are collected in lab medicine for ambulatory, inpatient reporting as well as ED. We asked our lab informatics to provide us with hemolysis rates and pulled frm our EMR-Epic potassium sample hemolysis rates

- **If Bonnie was <u>NOT</u> used, please identify the specifications for the test dataset** (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured)*
- **What levels of analysis were tested (either through the test data set or Bonnie)**? (*testing must be provided for <u>all</u> the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*) *in order to determine its suitability for inclusion into the Trial Approval Program.,*

Measure Specified to Measure Performance of:

Measure Tested at Level of:

☐ individual clinician

☐individual clinician

☐ group/practice

☐ group/practice

x hospital/facility/agency

☐ hospital/facility/agency

☐ other: Click here to describe

☐ other: Click here to describe

1.4. How many and which <u>patients</u> were included in the testing and analysis (by level of analysis and data source)? (*Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis)*

About 35,000/year ED patients who have a had lab potassium since 2014. For CDC funded project we have been looking at our EMR and Sunquest for all ED patients and have been using theis information to drive our performance improvement project. Most lab information systems(LIS) should be able to pull the data fields required for this project.

1.5. Please refer to the guidance for Bonnie testing found at this link. Bonnie testing results may be compiled into spreadsheet or table, which must be completed in its entirety, to the extent possible, in order to provide a basis for evaluation to determine the acceptability of the measure for inclusion in the Trial Approval program. *Any questions regarding the completion of this form can be directed to NQF Staff at trialmeasures@qualityforum.org.*

NA

<u>RELIABILITY AND VALIDITY ASSESSMENTS</u>

<u>Note</u>: The information provided in this next section is intended to aid the Standing Committee and other stakeholders in understanding to what degree the measure is both reliable and valid. While it is not possible to provide comprehensive results due to the lack of actual testing data, the developer needs to provide as much information as possible based on their interpretation of the results from the sample test data.

2.1 Reliability testing demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the

measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score. **What is your interpretation of the results in terms of demonstrating reliability**? (i.*e., what do the sample results mean and what are the norms for the test conducted?*) Please summarize the plan for future testing of reliability if the measure is accepted into the Trial Approval Program. Include descriptions of:

- Inter-abstractor reliability, and data element reliability of all critical data elements
- Computation of the performance measure score (e.g., signal-to-noise analysis)? There should be no

Validity against the Gold Standard will be assessed as follows.

EHR Measure Validity

The measure performance will be calculated from data collected using two different methods of collection:

- Automated EHR report
- Visual inspection of the medical record by professional data abstractors to capture the data elements to manually construct the performance

The data source and specific information about sites and patient records will be identified.

Data from a performance report for the measure automatically-generated from the EHR/LIS (designed to collect the necessary data elements to identify eligible cases and calculate the performance score) will be compared to data elements found and scores calculated manually on visual inspection of the medical record, by trained abstractors.

Data analysis will include:

- Percent agreement at the denominator, numerator
- Kappa statistic to ensure that agreement rates are not a phenomenon of chance

The selection of the testing method will depend on identification of test sites and availability of data.

2.2 Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score. **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?*). Please summarize the plan for future testing of validity if the measure is accepted into the Trial Approval Program. Include the method(s) of validity testing and what it will test (describe the steps—do not just name a method; what will be tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis will be used used)

Face validity of the measure score as an indicator of quality will be systematically assessed as follows.

The expert panel will be asked to rate their agreement with the following statement:

The scores obtained from the measure as specified will provide an accurate reflection of quality and can be used to distinguish good and poor quality.

Scale 1-5, where 1= Strongly Disagree; 3= Neither Agree nor Disagree; 5= Strongly Agree

The selection of the testing method will depend on identification of test sites and availability of data.

2.3 Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion. **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).* Please summarize the plan for future testing of exclusions if the measure is accepted into the Trial Approval Program. Describe the method of testing exclusions and what it will test (describe the steps—do not just name a method; what will be tested, e.g., whether exclusions affect overall performance scores; what statistical analysis will be used)

Not applicable

2.4 Risk Stratification (applicable ONLY to outcome or resource use measures). If an outcome or resource use measure will not be 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. If risk adjustment/stratification is needed then please describe the conceptual/clinical and statistical methods and criteria that will be used to select patient factors (clinical factors or sociodemographic factors) that will be 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*)

Not applicable

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

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

If other:

3b. Electronic Sources

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

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

ALL data elements are in defined fields in electronic health records (EHRs)

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.**

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.
Attachment **Attachment:** Blank_Feasibility_Assessment_Scorecard4fin.docx

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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
There are multiple ways to collect this data we collected data from both our ONC certified EMR Epic(Epic 14) and our ONC certified Laboratory information systems(LIS) (Sunquest 7.2). Obtaining data from our EMR was more difficult for us perhaps because we didn't purchase the lab add on(LIS) from Epic. The data required a few iterations before we felt we had the data we needed. Part of our difficulty was the ask was part of larger data ask revolving around a performance improvement project with need for other fields. Had we limited our request to just lab values and ED patients it may have been easier to obtain. Our organization chose to go with our current LIS for lab information Sunquest. Since this information can be obtained directly from the LIS we presumed this to be our reference value for ED hemolysis. Most labs have LIS that can extract this type of data and typically used for quality improvement projects. We presume a choice can be made for which system to submit from but most will go with their LIS initially because of the ease and familiarity. Preliminary data analysis from our EMR vs LIS(both Onc certified) for about 70,000 patients a year for 2 years with about 35,000 lab potassium results for ED patients showed hemolysis rates that were close but not an exact match likely do to data definitions and population definitions. We plan on submitted request for funding/grant to analyze the variance which we presume maybe more related to our data ask. For example our LIS vs EMR pulled 2015 data 22,892 vs 32, 327 patients, with gross hemolyzed 1.7% vs 2.1 while the hemolyzed with comment was 5.7% vs 6.9%. We are exploring if part of the reason was that the Sunquest data included information on our free standing ED's which the BI data may not have. Our plan is if we get funding either internally or from an EMF/ENA grant to include an analysis of why we had dropped patients from the BI/EMR side.

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)*.

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.

Planned	Current Use (for current use provide URL)
Public Reporting Public Health/Disease Surveillance Quality Improvement with Benchmarking	

(external benchmarking to multiple organizations) Quality Improvement (Internal to the specific organization)	

4a.1. **For each CURRENT use, checked above, provide:**

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

NA

4a.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?*)

ED hemolysis has significnat impact on care of ED patients. Since most of the issue aroudn this causes are preanaltycia; Its impact on ED patients and work in both ED and lab lead significantkly increased work around redraw and re-testing.

4a.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.*)

To be completed after testing of the measure

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**

Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

Organizations that have formally addressed the ED hemolysis problem (including Cleveland Clinic publications pending) have seen improvement in hemolysis rates by addressing some of the pre analytical issues some by either going to straight stick for most/all of ED blood samples or modification of equipment(use small volume/vaccum).

Heiligers-Duckers C, Peters NA, van Dijck JJ, Hoeijmakers JM, Janssen MJ. Low vacuum and discard tubes reduce hemolysis in samples drawn from intravenous catheters. Clin Biochem. 2013;46(12):1142-1144. Ong ME, Chan YH, Lim CS. Reducing blood sample hemolysis at a tertiary hospital emergency department. Am J Med. 2009;122(11):1054.e1-6.

4b.2. **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.**

One of the biggest reason is that very few ED are aware of the problem, nor are they aware that pre analytical factors(how the blood is drawn in the ED) are actually impacting hemolysis of ED samples. Many of the recommendations put forth have centered around use of straight needle to draw bloods necessitating the need for a second "stick" for IV placement but there are alternatives.

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**

None

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

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 completely harmonized?
No

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.)
NA

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:** testing_form_for_trial_useMay_6-2.docx

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Cleveland Clinic

Co.2 **Point of Contact:** Michael, Phelan, phelanm@ccf.org, 216-973-2003-
Co.3 **Measure Developer if different from Measure Steward:** Cleveland Clinic
Co.4 **Point of Contact:** Michael, Phelan, phelanm@ccf.org, 216-973-2003-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:**
Ad.3 **Month and Year of most recent revision:**
Ad.4 **What is your frequency for review/update of this measure?**
Ad.5 **When is the next scheduled review/update for this measure?**

Ad.6 **Copyright statement:**
Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 2988
Measure Title: Medication Reconciliation for Patients Receiving Care at Dialysis Facilities
Measure Steward: Kidney Care Quality Alliance (KCQA)
Brief Description of Measure: Percentage of patient-months for which medication reconciliation* was performed and documented by an eligible professional.**

* "Medication reconciliation" is defined as the process of creating the most accurate list of all home medications that the patient is taking, including name, indication, dosage, frequency, and route, by comparing the most recent medication list in the dialysis medical record to one or more external list(s) of medications obtained from a patient or caregiver (including patient-/caregiver-provided "brown bag" information), pharmacotherapy information network (e.g., Surescripts), hospital, or other provider.

** For the purposes of medication reconciliation, "eligible professional" is defined as: physician, RN, ARNP, PA, pharmacist, or pharmacy technician.
Developer Rationale: Medication management is a critical safety issue for all patients, but especially so for patients with ESRD, who often require 10 or more medications and take an average of 17-25 doses per day, have numerous comorbid conditions, have multiple healthcare providers and prescribers, and undergo frequent medication regimen changes(1,2,3,4). Medication-related problems (MRPs) contribute significantly to the approximately $40 billion in public and private funds spent annually on ESRD care in the United States(5,6), and it is believed that medication management practices focusing on medication documentation, review, and reconciliation could systematically identify and resolve MRPs, improve ESRD patient outcomes, and reduce total costs of care. As most hemodialysis patients are seen at least thrice weekly and peritoneal dialysis patients monthly, the dialysis facility has been suggested as a reasonable locale for medication therapy management(7).

Numerator Statement: Number of patient-months for which medication reconciliation was performed and documented by an eligible professional during the reporting period.

The medication reconciliation MUST:
• Include the name or other unique identifier of the eligible professional;

AND

• Include the date of the reconciliation;

AND

• Address ALL known home medications (prescriptions, over-the-counters, herbals, vitamin/mineral/dietary (nutritional) supplements, and medical marijuana);

AND

• Address for EACH home medication: Medication name(1), indication(2), dosage(2), frequency(2), route of administration(2), start and end date (if applicable)(2), discontinuation date (if applicable)(2), reason medication was stopped or discontinued (if applicable)(2), and identification of individual who authorized stoppage or discontinuation of medication (if applicable)(2);

AND

• List any allergies, intolerances, or adverse drug events experienced by the patient.

_____

1. For patients in a clinical trial, it is acknowledged that it may be unknown as to whether the patient is receiving the therapeutic agent or a placebo.

2. "Unknown" is an acceptable response for this field.

Denominator Statement: Total number of patient-months for all patients permanently assigned to a dialysis facility during the reporting period.

Denominator Exclusions: In-center patients who receive < 7 hemodialysis treatments in the facility during the reporting month.

Measure Type: Process
Data Source: Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record
Level of Analysis: Facility

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

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *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.

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 conducted a literature review which shows evidence to support the high incidence of medication-related problems in dialysis patients as well as evidence that supports their economic impact.

Guidance from the Evidence Algorithm

1-No→3-No→7-Yes→8-Yes→Moderate

Questions for the Committee:

For process 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?*

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

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

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

Performance scores over time are not available. However, the measure was tested using data from three KCQA member dialysis organizations, each with the capacity to provide retrospective analysis from a data warehouse repository. The study was conducted on data from April 1-September 30, 2015. Performance scores obtained during the testing are as follows:

Performance scores obtained during testing are as follows:

- Mean Performance Score = 52.62%
- Standard Deviation = 32.83
- Standard Error = 0.197
- 95% Confidence Interval = 52.24 to 53.01
- Median Score = 48.18
- Mode of Scores = 100
- Range of Scores = 0 to 100
- Interquartile Range = 27.59 to 87.62

Results show a significant spread between both the minimum and maximum scores, as well as the median and minimum and maximum scores, indicating there is significant room for improvement in this aspect of care and that the measure identifies clinically and practically meaningful differences in performance among the measured entities.

Disparities

- Empirical studies addressing medication management remain limited, and those focusing on dialysis patients or on sociodemographic discrepancies even more so. Two publications tangentially addressing such disparities among population groups were identified; only one was specific to the dialysis setting.
- One study reported a negative correlation between age and the number of drug record discrepancies identified ($r = -0.27$, $p = 0.04$) in hemodialysis patients.
- Another study found that medication discrepancies were more likely to persist in Caucasian subjects when compared to African Americans, despite pharmacist-led medication reconciliation.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this*

area of healthcare?

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)

1. Importance to Measure and Report

1a. Evidence to Support Measure Focus

Comments: **This is a process measure. A systematic review of the evidence specific to this measure was not completed. The developer did include some evidence of the importance of medication reconciliation in reducing medication related problems in ESRD patients. However, several of the studies referenced medication reconciliation, review and management - not reconciliation alone - as the effective intervention. In addition, the gold standard appears to be reconciliation performed by a pharmacist and this measure allows for other clinicians. If the goal is to reduce MRPs in ESRD patients, the evidence would suggest that this measure, as specified, may fall short of the desired outcome.
**This is a process measure. Medication reconciliation in this high risk (in a complicated state of pseudo-stability) and who take lots of medications is important and does link to patient safety.
**A literature review was done that shows evidence to support the high incidence of med-related problems.

1b. Performance Gap

Comments: **The developer tested the measure in three dialysis organizations using large sample of patients. The resulting performance scores do indicate a gap in performance. The mean performance score was 52.62% with an interquartile range of 27.59 to 87.62.
**That the reported rate is only 52% shows in itself a big gap. This gap may be related to the level of detail required for this measure medication reconciliation (all elements on all medications and adverse drug reactions). This is no longer just an attestation. This measure has more "meat" to it. In addition, that this would be required 1/month highlights even more so the performance gap
**"Measure was tested using retrospective data (April 1 - Sept. 30, 2015) from 3 Kidney Care Quality Alliance members Results showed a significant spread between minimum and maximum scores, median and minimum and maximum scores. Clinically and practically meaningful differences are possible.There is very little data on medication management for dialysis patients or on SDS discrepancies. One study showed a negative correlation between age and the number of drug record discrepancies identified in hemodialysis patients."

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record

Numerator Statement: Number of patient-months for which medication reconciliation was performed and documented by an eligible professional during the reporting period.

The medication reconciliation MUST:
• Include the name or other unique identifier of the eligible professional;
AND

• Include the date of the reconciliation;
AND
• Address ALL known home medications (prescriptions, over-the-counters, herbals, vitamin/mineral/dietary (nutritional) supplements, and medical marijuana);
AND
• Address for EACH home medication: Medication name(1), indication(2), dosage(2), frequency(2), route of administration(2), start and end date (if applicable)(2), discontinuation date (if applicable)(2), reason medication was stopped or discontinued (if applicable)(2), and identification of individual who authorized stoppage or discontinuation of medication (if applicable)(2);
AND
• List any allergies, intolerances, or adverse drug events experienced by the patient.

NUMERATOR STEP 1. For each patient meeting the denominator criteria in the given calculation month, identify all patients with each of the following three numerator criteria (a, b, and c) documented in the facility medical record to define the numerator for that month:

A. Facility attestation that during the calculation month:

1. The patient's most recent medication list in the dialysis medical record was reconciled to one or more external list(s) of medications obtained from the patient/caregiver (including patient-/caregiver-provided "brown-bag" information), pharmacotherapy information network (e.g., Surescripts®), hospital, or other provider AND that ALL known medications (prescriptions, OTCs, herbals, vitamin/mineral/dietary [nutritional] supplements, and medical marijuana) were reconciled;

AND

2. ALL of the following items were addressed for EACH identified medication:
a) Medication name;
b) Indication (or "unknown");
c) Dosage (or "unknown");
d)Frequency (or "unknown");
e) Route of administration (or "unknown");
f) Start date (or "unknown");
g) End date, if applicable (or "unknown");
h) Discontinuation date, if applicable (or "unknown");
i) Reason medication was stopped or discontinued, if applicable (or "unknown"); and
j) Identification of individual who authorized stoppage or discontinuation of medication, if applicable (or "unknown");

AND

3. Allergies, intolerances, and adverse drug events were addressed and documented.

B. Date of the medication reconciliation.

C. Identity of eligible professional performing the medication reconciliation.

NUMERATOR STEP 2. Repeat "Numerator Step 1" for each month of the one-year reporting period

to define the final numerator (patient-months).

Denominator Statement: Total number of patient-months for all patients permanently assigned to a dialysis facility during the reporting period.

Denominator Exclusions: In-center patients who receive < 7 hemodialysis treatments in the facility during the reporting month.

DENOMINATOR STEP 1. Identify all in-center and home hemodialysis and peritoneal dialysis patients permanently assigned to the dialysis facility in the given calculation month.

DENOMINATOR STEP 2. For all patients included in the denominator in the given calculation month in "Denominator Step 1", identify and remove all in-center hemodialysis patients who received < 7 dialysis treatments in the calculation month.

DENOMINATOR STEP 3. Repeat "Denominator Step 1" and "Denominator Step 2" for each month of the one-year reporting period.

Questions for the Committee :

- *Are all the data elements clearly defined?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing **Testing attachment**

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

N/A

SUMMARY OF TESTING

Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure

☒ **Yes** ☐ **No**

Method(s) of reliability testing Beta-binomial testing

Results of reliability testing

For the 6-month study period, for all facilities (excluding those with <=11 patients in a given reporting month, as per the measure specifications [approximately 3.7% of facilities each month]), the mean reliability of the measure is 0.9935 (range = 0.8166-1).

Guidance from the Reliability Algorithm

There was empirical reliability testing conducted using statistical tests, specifically at the computed measure score. There was an appropriate statistical test used. Based on the results of the reliability testing, there is a high level of certainty that the performance scores are reliable.

Questions for the Committee:

- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

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

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ Yes ☐ Somewhat ☐ No

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

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, summarize the validity testing from the prior review:

N/A

SUMMARY OF TESTING
Validity testing level ☒ Measure score ☐ Data element testing against a gold standard ☐ Both

Method of validity testing of the measure score:

- ☒ **Face validity only**
- ☐ **Empirical validity testing of the measure score**

Validity testing method: There was a systematic assessment of face validity by experts. Two groups of field experts in the field of ESRD / dialysis care. Each group completed a face validity assessment that explicitly addressed whethe performance scores resulting from the measure, as specified, provide an accurate reflection of quality. Individuals responded to the following two questions:

1. How likely is it that the measure score provides an accurate reflection ofmedication reconciliation quality? (highly unlikely; unlikely; neither likely nor unlikely; likely; highly likely)

2. What is the likelihood that the measure can be used to distinguish good from poor quality? (highly unlikely; unlikely; neither likely nor unlikely; likely; highly likely)

Validity testing results:

The face validity assessment yielded the following:
• KCQA Member Organizations' Lead Representatives:

77.3% of KCQA Lead Representatives (n=22) agreed it is highly likely or likely that the measure score provides an accurate reflection of medication reconciliation quality.

77.3% of the panel agreed that it is likely/highly likely that the measure can be used to distinguish good from poor quality.

• Expert Panel:

88.9% of the 9-member panel agreed it is highly likely or likely that the measure score provides an accurate reflection of medication reconciliation quality.

77.8% of the panel agreed it is highly likely or likely that the measure can be used to distinguish good from poor quality.

Questions for the Committee:

- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

There is one patient level and one facility level exclusion:

Patient-Level Exclusion: Transient patients, i.e., in-center hemodialysis patients who receive <7 dialysis treatments in the facility during the calculation month.

Facility-Level Exclusion: Facilities with <=11 (i.e., <12) patients during the calculation month.

Based on the results presented by the measure developer, there were less than 2 patients excluded per facility per study month. A total of 2.8% of patients on average were excluded. Based on the facility-level exclusion, approximately 3.7% of facilities were excluded.

Based on these results, the developer concluded that the variability and frequency with which exclusions were encountered during tests was sufficient to demonstrate they are necessary to prevent unfair distortion of the results.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

The developer provided descriptive statistics of the measure:

Mean Performance Score = 52.62%
- Standard Deviation = 32.83
- Standard Error = 0.197
- 95% Confidence Interval = 52.24 to 53.01
- Median Score = 48.18
- Mode of Scores = 100
- Range of Scores = 0 to 100
- Interquartile Range = 27.59 to 87.62

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

N/A

2b7. Missing Data

There were no missing data to report.

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

Committee pre-evaluation comments

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

2. Scientific Acceptability of Measure Properties

2a1. & 2b1. Specifications

Comments: **"The measure is reported as the percent of patient-months for which medication reconciliation was performed and documented by an eligible professional. The measure only includes patients who were permanently assigned to the dialysis center. I'm confused about why the measure is collected and reported as patient-months instead of patients. I'm also confused on how to calculate the numerator. Do you count the month as ""1"" if all patients that month met all three criteria (A-C)? Or do you want the actual patients within that month where all three criteria (A-C) were met? In other words, if one case falls out of a particular month (i.e. failed to meet the criteria A-C) does the month then fall out of the denominator. I find the concept of patient-months confusing and feel that patients will as well.

**While the specifications for medication reconciliation are consistent, the specifications as to what constitutes "adverse drug reaction" are not spelled out in detail. Perhaps the developers mean "patient reported Adverse drug reaction" as opposed to patient "experienced" (this implies the experience is easily determined. The developers have to provide more detail to what is an Adverse Drug Reaction. Also it is unclear as to whether peritoneal dialysis is also included. what is also unclear is to when a numerator is not "perfect" - is it when only one element is missing or off? What is also missing is inter-rater reliability testing as to the patient interview process whereas the elements are captured.

**"Specifications are clear. Specifications are consistent with evidence.

2a2. Reliability Testing

Comments: **The measure was tested at the performance score level with a mean reliability score of 0.9935 indicating extremely high reliability.

**The reported reliability is 0.7 (okay) however the developers do state that there will be required upgrading and programing of any electronic medical record to accomplish this measure detail. Going to an EMR is great but relying on "paper record" and inconsistent means to determine current medications makes me surprised that the reliability is that high. Patient reported medications have a reported error in themselves of up to 50-60% (not from the developers).

**Measure score reliability testing was done using a beta-binomial model. The mean reliability was 0.9935, with a range of 0.8166 to 1.0. High reliability

2b2. Validity Testing

Comments: **The measure underwent testing for face validity only by two groups of ESRD experts: KCQA member organization lead representatives and an expert panel. Both groups rated the measure as likely or highly likely to accurately reflect med rec quality and distinguish between good and poor quality.
**Validity was not tested or reported and the authors do not provide any data that where medication reconciliation picks up adverse drug reactions or errors that have/can cause harm
**Measure score validity testing was done using face validity. Two groups of experts completed the assessment.77.3% of KCQA experts agreed that the measure score provides an accurate reflection of medication reconciliation quality and that the measure can be used to distinguish good from poor quality. 88.9% of the second expert panel agreed that the measure provides an accurate reflection of med rec quality; 77.8% agreed that the measure can distinguish good from poor quality.

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **There is no explanation of how the exclusion criteria were selected. patients receiving <7 treatments at the facility during the treatment month are excluded. Facilities with <12 patients during the calculation month are excluded.
**Exclusions are spelled out but I see no evidence as to why dialysis centers with less than 11 patients are excluded. Though the developers report this represents only 3.7% of facilities, this may be a larger percent for hospitals. I believe this number has to be reviewed again
**Exclusions are supported by the analysis. The descriptive statistics indicate this measure identifies meaningful differences in quality.

Criterion 3. Feasibility

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.

- This measure is generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score)
- ALL data elements are in defined fields in electronic health records (EHRs).
- MEDICATION MANAGEMENT DEFINITIONAL DISCREPANCIES were identified when developing the measure specifications and operationalizing the specifications for testing, it was noted while all three dialysis organizations that participated in testing have identified and engage in the same three components of medication management—i.e., documentation, reconciliation, and review—one organization defined reconciliation and review in reverse to those detailed in the measure specifications.
- DATA SYSTEM DISCREPANCIES were identified when developing the measure specifications and operationalizing the specifications for testing, variations between the electronic medical record systems of the three large dialysis organizations that participated in testing were identified. For instance, a given data element (e.g., indication, start date, name of eligible professional) might not be present or might be available only as a free text field. It was further noted that this variability might be even greater in the medium and small dialysis organizations. Given the variability among electronic systems and because some medications are prescribed by other entities for which "indication" may be unknown, for example, it was determined that "unknown" must be an allowable response to many data elements so as to maintain the measure's feasibility.
- No fees or licensing requirements to use any aspect of the measure as specified, were reported.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*

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

Committee pre-evaluation comments

Criteria 3: Feasibility

3. Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **Two issues were identified during testing: medication management definitional discrepancies and data system discrepancies. Neither seemed to impact the reliability of the measure. However, it was noted that issues may arise if the measure is rolled out to medical and small dialysis centers.
**In its current state, the measure does not appear to be "ready" as there are EMR changes that must occur as well as programming to capture the elements and lack of reconciliation. In addition, there needs to be more definition/education around Adverse Drug Reactions.
**Data for the measure is generated during care provision. All data elements are in an EHR. There were discrepancies in the definition of medication reconciliation identified between dialysis organizations. Data system discrepancies were also identified in the medical record systems of 3 dialysis organizations. Indications for use of medications might not be present in some systems; so "unknown" has to be included as an allowed response.

Criterion 4: Usability and Use

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

Current uses of the measure:

Publicly reported? ☐ **Yes** ☒ **No**

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

OR

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

Accountability program details:

- This is a new measure that is not yet in use as specified.
- Variants of the measure are currently in use member dialysis organizations for internal quality improvement, prompting the developer to develop this measure to standardize the specifications and definitions for accountability purposes.
- Panned use includes: Public Reporting, Payment Program, Payment Program, Quality Improvement with Benchmarking (external benchmarking to multiple organizations), and

Quality Improvement (Internal to the specific organization).

- The measure was developed for use by CMS for its accountability initiatives. The developer notes that the measure requires a number of data fields not currently available in the CROWNWeb ESRD clinical data repository, and would require a system update for implementation.

Improvement results:

- There were no unintended consequences were identified during testing.

Potential harms:

- No unintended consequences were identified during testing.

Feedback :

- Developer did not identify any specific feedback loops related to this 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

4. Usability and Use

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **"This is a new measure. The measure is not currently in use. The measure was developed to be used by CMS for public reporting and payment. The developer notes that a number of data fields are not currently available in the CROWNWeb ESRD clinical data repository and would require a system update for implementation.

**While I believe this measure really raises the bar to quality and safety and is important. I believe it would be burdensome in its current state and more specific as to what constitutes a reconciliation data piece error "address"

**The measure is not publicly reported and is not used in an accountability program. Variants of the measure are used in member dialysis orgs for internal QI. CMS intends to use this measure in its accountability initiatives.

Criterion 5: Related and Competing Measures

Related or competing measures

- 0097 : Medication Reconciliation Post-Discharge
- 0554 : Medication Reconciliation Post-Discharge (MRP)
- 2456 : Medication Reconciliation: Number of Unintentional Medication Discrepancies per Patient

Harmonization

- This measure is harmonized with existing NQF-endorsed medication reconciliation measures in that all similarly specify that the medication reconciliation must address ALL prescriptions, over-the-counters, herbals, vitamin/mineral/dietary (nutritional) supplements AND must contain the medications' name, dosage, frequency, and route. This measure, however, is unique among the currently endorsed medication reconciliation measures in that the level of analysis is the dialysis facility. The KCQA measure also moves beyond a single "check/box", specifying multiple components that must be met to be counted as a "success".

Pre-meeting public and member comments

-

National Quality Forum—Evidence (subcriterion 1a)

Measure Number: NQF 2988
Measure Title: ***Medication Reconciliation for Patients Receiving Care at Dialysis Facilities.*** **IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here:** Not applicable.
Date of Submission: 5/10/2016

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins).

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status,

symptom/symptom burden, experience with care, health-related behavior.

- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

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

Outcome

- □ Health outcome: Click here to name the health outcome

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health--- related behaviors

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

☒ Process: **Monthly medication reconciliation for patients receiving care at dialysis facilities.**

- □ Structure: Click here to name the structure
- □ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

Implementation of standardized medication reconciliation definitions, specifications, and frequency for accountability purposes by dialysis facilities

↓

Medication reconciliation (the process of creating the most accurate list of all home medications that the patient is taking, including name, indication, dosage, frequency, and route, by comparing the most recent medication list in the dialysis medical record to one or more external list(s) of medications obtained from a patient or caregiver, pharmacotherapy information network, hospital, or other provider) performed on a monthly basis for all dialysis patients

↓

Improved and expedited identification of real and potential medication---related problems (MRPs) in ESRD patients

↓

Reduction of MRP---associated hospitalizations, readmissions, mortality, and health care costs

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

- ☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***
- ☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***
- ☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☒ **Other – *complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence---based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

- ☐ Yes → ***complete section 1a.7***
- ☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE
1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE
If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality,

and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990---2010*). Date range: Click here to enter date range

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta---analysis, and statistical significance*)

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for each new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

1a.8 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.8.1 What process was used to identify the evidence? A search of relevant literature was conducted to identify studies on outcomes of medication reconciliation.

1a.8.2. Provide the citation and summary for each piece of evidence. Citations are listed alphabetically by lead author. We note 5 of the 11 citations are not specific to the dialysis population,

but given the previously---noted existing lack of empirical publications specifically addressing the ESRD population, these provide applicable background information on current medication management practices and medication---related problems.

ESRD POPULATION---SPECIFIC CITATIONS:

1. **Cardone KE, Bacchus S, Assimon MM, Pai AB, Manley HJ. Medication---related problems in CKD. *Adv Chronic Kidney Dis.* 2010;17(5):404---412.**

Summary: The authors note patients with CKD often are prescribed heterogeneous medications to treat disease---associated comorbidities, to slow down progression of the disease, and to minimize morbidity and mortality rates. Medication regimens in this population are very complex, leading to an increased potential for medication---related problems (MRPs). As kidney function declines, the type and amount of medications a patient consumes increases, thereby putting CKD patients at a higher risk for MRPs. MRPs have been known to be associated with morbidity, mortality, and a lower quality of life.

2. **Hakim RM, Collins AJ. Reducing avoidable rehospitalization in ESRD: A shared accountability. *JASN.* 2014;25(9):1891---1893.**

Summary: The authors note interventions and services by the healthcare team that can lead to reduced rehospitalization include one or more episodes of medication reconciliation facilitated by a knowledgeable pharmacist in the dialysis facility after each rehospitalization, and conclude medication reconciliation after hospital discharge is critically needed because it crosses all aspects of care. The article indicates ESRD patients are prescribed an average of 11---12 medications and take an average of 17---25 doses per day, and thus experience a high rate of medication---related problems (MRPs). MRPs are particularly acute at the time of hospital discharge because that process often involves changes to the prehospitalization prescribed medications. The involvement of pharmacists has been shown to both identify actual and potential MRPs, as well as to reduce rehospitalizations and lengths of stay of dialysis patients.

3. **Manley HJ, Carroll CA. The clinical and economic impact of pharmaceutical care in end---stage renal disease patients. *Semin Dial.* 2002;15:45---49.**

Summary: The authors note ESRD patients are medically complex, require multiple medications for treatments of their various comorbidities, and cost the healthcare system billions of dollars each year. These patients are also at risk of drug---related problems (DRPs) that may lead to increased morbidity, mortality, and cost to the healthcare system. The authors note the literature demonstrates pharmaceutical care provided by pharmacists improves ESRD patient care. Specifically, pharmacist review of ESRD patients' medication profiles and medical records has shown to be beneficial in identifying and resolving DRPs, and an economic analysis suggests that for every $1 spent on pharmaceutical care, the health care system saves an estimated $3.98. The authors conclude provision of pharmaceutical care by pharmacists should be considered for all ESRD patients.

4. **Manley HJ, Drayer DK, McClaran M, Bender W, Muther RS. Drug record discrepancies in an outpatient electronic medical record: Frequency, type, and potential impact on patient care at a hemodialysis center. *Pharmacotherapy.* 2003;23(2):231---239.**

Summary: The authors noted electronic drug record discrepancies are a potential source of drug--- related problems and sought to determine the extent to which such discrepancies occur in a hemodialysis population through a prospective observational study of patients enrolled in a

pharmacist clinic at an outpatient hemodialysis center from August---December 2001. Patients participated in monthly drug interviews conducted by a pharmacist, during which drug record discrepancies were classified and assigned a potential drug---related problem. Patients with documented drug record discrepancies were compared with those patients for whom no discrepancy was identified. Over the 5---month period, 215 drug interviews were conducted for 63 patients; 113 drug record discrepancies were identified in 38 patients (60%). Electronic drug records were discrepant by one, two, and more than two drug records 60.0%, 26.2%, and 13.8% of the time, respectively. Drug record discrepancies placed patients at risk for adverse drug events and dosing errors in 49.6% and 34.5%, respectively, of 113 discrepancies. Patient age negatively correlated with the number of drug record discrepancies identified (r=---0.27, p=0.04). The authors concluded drug record discrepancies occur frequently among hemodialysis patients, and that incorporation of a pharmacist into the patient care team may increase the accuracy of the electronic drug records and avert unnecessary drug---related problems.

5. **Pai AB, Boyd A, Depczynski J, Chavez IM, Khan N, Manley H. Reduced drug use and hospitalization rates in patients undergoing hemodialysis who received pharmaceutical care: A 2---year, randomized, controlled study. *Pharmacotherapy.* 2009; 29: 1433–1440.**
   Summary: The authors conducted a prospective, randomized, controlled, longitudinal, 2---year pilot study intended to investigate the impact of a pharmaceutical care program managed by clinical pharmacists on drug use, drug costs, hospitalization rates, and drug---related problems (DRPs) in 104 adult ambulatory patients undergoing hemodialysis in a nonprofit university---affiliated dialysis clinic. Patients were randomly assigned to receive either pharmaceutical care, consisting of one---on---one care with in---depth drug therapy reviews conducted by a clinical pharmacist (57 patients), or standard of care, consisting of brief drug therapy reviews conducted by a nurse (47 patients). Baseline data on demographic and clinical characteristics were collected, and mean numbers of concomitant drugs, drug costs, hospitalization rates, and lengths of stay were compared between the groups. In the pharmaceutical care group, DRPs were identified and recorded. Baseline age, length of time receiving hemodialysis, etiology of ESRD, and mean number of concomitant drugs at baseline were similar between the groups.

At the end of the 2---year follow---up, the authors found pharmaceutical care was associated with a significant decrease of 14% fewer drugs compared with standard of care, as documented during each drug therapy review ($p<0.05$). There were significantly fewer all--- cause hospitalizations among patients assigned to pharmaceutical care compared with those receiving standard of care (mean +/--- SD 1.8 +/--- 2.4 vs 3.1 +/--- 3 hospitalizations, p=0.02), and the cumulative time hospitalized was shorter in the pharmaceutical care group compared with the standard of care group (9.7 +/--- 14.7 vs 15.5 +/--- 16.3 days, p=0.06). During the study period, 530 DRPs were identified and resolved. The authors concluded the provision of pharmaceutical care is associated with tangible benefits on outcomes in ambulatory patients undergoing hemodialysis and should be considered in health care policy decisions.

6. **Spiegel B, Bolus R, Desai AA, Zagar P, Parker T, Moran J, Solomon MD, Khawar O, Gitlin M, Talley J, Nissenson A. Dialysis practices that distinguish facilities with below--- versus above---expected mortality. *CJASN.* 2010;5:2024---2033.**

Summary: The authors noted mortality rates vary widely among dialysis facilities, even after adjustment with standardized mortality ratios (SMRs); they hypothesized this

variation may occur because either top---performing facilities use practices not shared by others, the SMR fails to capture key patient characteristics, or both. The authors identified specific practices, including frequency of medication reconciliation by nurses, that distinguish top--- from bottom---performing facilities by SMR. A cross---sectional survey of staff was performed across three dialysis organizations. Staff members rated the perceived quality of their units' patient---, provider---, and facility---level practices using a six--- point Likert scale. Facilities were divided into those with above--- versus below---expected mortality on the basis of SMRs from U.S. Renal Data Service facility reports. Mean Likert scores were computed for each practice using t tests. Practices that were statistically significant ($P \leq 0.05$) and achieved at least a medium effect size of ≥0.4 were reported.
Significant predictors were entered into a linear regression model.

Dialysis facilities with below---expected mortality reported that patients in their unit were more activated and engaged, physician communication and interpersonal relationships were stronger, dieticians were more resourceful and knowledgeable, and overall coordination and staff management were superior versus facilities with above---expected mortality. Importantly, units with lower---than---expected mortality rates engaged in a more coordinated, multidisciplinary environment, including (but not limited to) convening multidisciplinary conferences sooner after dialysis patients return to the facility after hospitalization and performing medication reconciliation more frequently than high--- mortality units. Staff ratings of these practices explained 31% of the variance in SMRs.

GENERAL POPULATION CITATIONS:

7. **Bedell SF, Jabbour S, Goldberg R et al. Discrepancies in the use of medications: Their extent and predictors in an outpatient practice. *Arch Intern Med.* 2000;160:2129---2134.**
   Summary: The authors noted misuse of medications is a major cause of morbidity and mortality, and few studies had yet examined the frequency of and factors associated with discrepancies between what doctors prescribe and what patients take in actual practice. Specifically, 312 patients from the practices of 5 cardiologists and 2 internists who were returning for their routine follow---up visits were included in the study. Patients' medication bottles and their reported use of medications were compared with physicians' records of outpatients seen between November 1997 and February 1998 in a private practice affiliated with an academic medical center in Boston, MA. Discrepancies were found in medications for 239 patients (76%). The 545 discrepancies were the result of patients taking medications that were not recorded (n = 278 [51%]), patients not taking a recorded medication (n = 158 [29%]), and differences in dosage (n = 109 [20%]). Overall, discrepancies were randomly distributed among different drugs and discrepancy types with no discernible pattern. Multivariate analysis revealed patient age and number of recorded medications were the 2 most significant predictors of medication discrepancy.

Discrepancies among recorded and reported medications were common and involved all classes of medications, including cardiac and prescription drugs. Older age and polypharmacy were the most significant correlates of discrepancy. The authors concluded the pervasiveness of discrepancies can have significant health care implications, and action is urgently needed to address their causes; such action would likely have a positive impact on patient care.

8. **Isetts BJ, Schondelmeyer SW, Artz MB, Lenarz LA, Heaton AH, Wadd WB, Brown LM,**

Cipolle RJ. Clinical and economic outcomes of medication therapy management services: The Minnesota experience. *J Am Pharm Assoc.* 2008;48:203–211.

Summary: The authors conducted a prospective study of six ambulatory clinics in Minnesota from August 1, 2001, to July 31, 2002 consisting of 285 intervention group patients with at least 1 of 12 medical conditions using pre---study health claims, 126 comparison group patients with hypertension, and 126 patients with hyperlipidemia selected among 9 clinics without Medication Therapy Management (MTM) services for HEDIS analysis. The authors assessed the clinical effects associated with the provision of MTM services by measuring the percent of patients achieving HEDIS goals for hypertension and hyperlipidemia in the MTM services intervention group in relationship to a comparison group who did not receive MTM services. Patients' total health expenditures for the year before and after receiving MTM services were also compared. MTM services were provided by pharmacists to health plan beneficiaries in collaboration with primary care providers. Main outcomes included resolution of drug therapy problems, percentage of patients' goals of therapy achieved, and meeting HEDIS measures for hypertension and hypercholesterolemia. Total health expenditures per person were measured for a 1---year period before and after enrolling patients in MTM services.

Findings from the study were: 637 drug therapy problems were resolved among 285 intervention patients, and the percentage of patients' goals of therapy achieved increased from 76% to 90%. HEDIS measures improved in the intervention group compared with the comparison group for hypertension (71% versus 59%) and cholesterol management (52% versus 30%). Total health expenditures decreased from $11,965 to $8,197 per person (n = 186, P < 0.0001). The reduction in total annual health expenditures exceeded the cost of providing MTM services by more than 12 to 1. The authors concluded patients receiving face---to---face MTM services provided by pharmacists in collaboration with prescribers experienced improved clinical outcomes and lower total health expenditures. Clinical outcomes of MTM services have chronic care improvement and value---based purchasing implications, and economic outcomes support inclusion of MTM services in health plan design.

9. **Stewart AL, Lynch KJ. Medication discrepancies despite pharmacist led medication reconciliation: The challenges of maintaining an accurate medication list in primary care. *Pharm Pract.* 2014;12(1)360.**

Summary: The authors report on an observational case series study of established patients from an urban, indigent care clinic intended to describe the types of medication discrepancies that persist despite pharmacist---led medication reconciliation using the primary care electronic medical record (EMR). Medication reconciliation was conducted immediately prior to the physician visit at baseline and return visit. Main outcome measures included frequency, types, and reasons for discrepancies, patient knowledge, and adherence. There was a 14.5% reduction in the number of patients with a discrepancy, the frequency of discrepancies was reduced by 7.3%, and the rate of medication discrepancies in the chart was reduced by 31.3% with pharmacist---led medication reconciliation. The most common type of discrepancy that persisted at follow--- up despite the intervention were medications listed on the chart that the patient had discontinued. Additionally, discrepancies were more likely to persist despite the pharmacist---led intervention in Caucasian subjects when compared to African Americans. The authors concluded that while pharmacist led medication reconciliation appears

effective at reducing the likelihood of a medication discrepancy in the EMR, challenges persist in maintaining this accuracy, specifically as it relates to patient---driven changes to the medication regimen.

10. **Tache SV, Sonnichsen A, Ashcrof, DM. Prevalence of adverse drug events in ambulatory care: A systematic review. *The Annals of Pharmacotherapy*. 2001; 45(7---8):977---989.**

Summary: The authors note while most medications are prescribed, dispensed, and administered in ambulatory care settings, little information exists on the adverse effects of drugs in this setting. This review was conducted to estimate the prevalence of adverse drug events (ADEs) and the proportion of preventable ADEs in ambulatory care settings, as well as to compare data for different age groups and review drug classes most commonly associated with ADEs. Four electronic databases—PubMed (1966---March 2011), International Pharmaceutical Abstracts (1970---March 2011), EMBASE (1980---March 2011), and the Cochrane Database of Systematic Reviews (1993---March 2011)—were systematically searched for published data, and bibliographies of retrieved articles were searched individually for additional relevant studies. A standardized definition of ADE was used to select studies in populations living in the community, with medical visits to primary care facilities, non---specialty ambulatory care facilities, and/or admissions to a hospital for medication---related adverse events. Forty---three studies met inclusion criteria.

The median ADE prevalence rate for retrospective studies was 3.3% (interquartile range [IQR] 2.3---7.1%) vs 9.65% (IQR 3.3---17.35%) for prospective studies. Median preventable ADE rates in ambulatory care---based studies were 16.5%, and 52.9% for hospital---based studies. Median prevalence rates by age group ranged from 2.45% for children to 5.27% for adults, 16.1% for elderly patients, and 3.45% for studies including all ages. The authors concluded the identified notable differences in prevalence rates by age groups and responsible drug categories offer guidance on how to direct attention toward effective targets for improvement of medication safety in ambulatory care settings.

11. **Wagner MM, Hogan WR. The accuracy of medication data in an outpatient electronic medical record. *J Am Med Inform Assoc.* 1996;3:61---68.**

Summary: The objective of this prospective cohort study was to measure the accuracy of medication records stored in the electronic medical record (EMR) of an outpatient geriatric center. The authors analyzed accuracy from the perspectives of a clinician using the data and a computer--- based medical decision---support system (MDSS). During scheduled office visits for medical care, the treating clinician determined whether the medication records for the patient were an accurate representation of the medications the patient was actually taking. Using the available sources of information (the patient, the patient's vials, any caregivers, and the medical chart), the clinician determined whether the recorded data were correct, whether any data were missing, and the type and cause for each discrepancy found.

The authors found 83% of medication records correctly represented the compound, dose, and schedule of a current medication; 91% represented correctly the compound; and 0.37 current medications were missing per patient. The principal cause of errors was found to be the patient (36.1% of errors), who misreported a medication at a previous visit or changed (stopped, started, or dose---adjusted) a medication between visits. The second most frequent cause of errors was failure to capture changes to medications made by outside clinicians, accounting for 25.9% of errors.
Transcription errors comprised 8.2% of errors. When the accuracy of records from the center was

analyzed from the perspective of an MDSS, 90% were correct for compound identity and 1.38 medications per patient were missing or uncoded. The cause of the additional errors of omission was a free---text "comments" field, assumed to be unreadable by current MDSS applications, used by clinicians in 18% of cases to record the identity of the medication. The authors concluded medication records in an outpatient EMR may have significant levels of data error. Based on an analysis of correctable causes of error, the authors suggested the most effective extension to the EMR studied would be to expand its scope to include all clinicians who can potentially change medications. However, even with EMR extensions ineradicable error due to patients and data entry will likely remain. It was noted the provision of a free---text "comments" field increased the accuracy of medication lists for clinician users at the expense of accuracy for an MDSS.

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
MM-2_NQF_EvidenceAttachment05-10-16FINAL.pdf

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., the benefits or improvements in quality envisioned by use of this measure)*

Medication management is a critical safety issue for all patients, but especially so for patients with ESRD, who often require 10 or more medications and take an average of 17-25 doses per day, have numerous comorbid conditions, have multiple healthcare providers and prescribers, and undergo frequent medication regimen changes(1,2,3,4). Medication-related problems (MRPs) contribute significantly to the approximately $40 billion in public and private funds spent annually on ESRD care in the United States(5,6), and it is believed that medication management practices focusing on medication documentation, review, and reconciliation could systematically identify and resolve MRPs, improve ESRD patient outcomes, and reduce total costs of care. As most hemodialysis patients are seen at least thrice weekly and peritoneal dialysis patients monthly, the dialysis facility has been suggested as a reasonable locale for medication therapy management(7).

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

Medication Reconciliation for Patients Receiving Care at Dialysis Facilities is new measure that is not yet in use, so performance scores over time are not available. However, the measure was tested using data from three KCQA member dialysis organizations, each with the capacity to provide retrospective analyses from a data warehouse/repository. All pertinent data from all eligible (i.e., adult and pediatric in-center and home hemodialysis and peritoneal dialysis) patients of the participating organizations during the testing period were included in the dataset. The number of patients and contributing facilities varied by month, but approximately 325,000 patients and 5,292 facilities across the three organizations were included in each of the six months of the study. The study was conducted on data from April 1-September 30, 2015.

Performance scores obtained during testing are as follows:

- Mean Performance Score = 52.62%
- Standard Deviation = 32.83

- Standard Error = 0.197
- 95% Confidence Interval = 52.24 to 53.01
- Median Score = 48.18
- Mode of Scores = 100
- Range of Scores = 0 to 100
- Interquartile Range = 27.59 to 87.62

Results show a significant spread between both the minimum and maximum scores, as well as the median and minimum and maximum scores, indicating there is significant room for improvement in this aspect of care and that the measure identifies clinically and practically meaningful differences in performance among the measured entities.

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.**
Testing data are presented in 1b.2. Contemporary literature supports our findings documenting variations in performance and room for improvement in medication management practices in dialysis facilities.

As previously noted, ESRD patients often are prescribed 10 or more medications, have multiple comorbidities and numerous healthcare providers, and undergo frequent medication regimen changes, putting them at high risk for medication errors, discrepancies, and other medication-related problems (MRPs)(1,2,3,4). While there is a paucity of peer-reviewed empirical studies addressing medication management specifically in dialysis facilities, those that have been published provide convincing evidence for the need for increased focus in this area.

One small prospective observational study (2003) in a single outpatient hemodialysis center identified discrepancies in 60% of participating patients' home medications lists when compared to those documented in the dialysis facility medical record(8). A 2009 randomized controlled trial demonstrated an association between increased focus on medication management in dialysis facilities and the identification of real and potential MRPs, as well as a decrease in the numbers of drugs taken by ESRD patients and a reduction in all-cause hospitalization rates and hospital lengths-of-stay(1,9,10). Likewise, the Identifying Best Practices in Dialysis (IBPiD) Study, a cross-sectional staff survey of three dialysis organizations comparing the perceived quality of patient-, provider-, and facility-level practices with Standardized Mortality Ratio (SMR) scores from U.S. Renal Data Service (USRDS) facility reports, found units with lower-than-expected mortality rates convene multidisciplinary conferences sooner after dialysis patients return to the facility after hospitalization and perform medication reconciliation more frequently than high-mortality units(11).

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*
Not applicable—new measure; not yet in use.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
Again, empirical studies addressing medication management remain limited, and those focusing on dialysis patients or on sociodemographic discrepancies even more so. Two publications tangentially addressing such disparities among population groups were identified; only one was specific to the dialysis setting. Specifically, the previously mentioned 2003 observational study by Manley et al. reported a negative correlation between age and the number of drug record discrepancies identified ($r = -0.27$, $p = 0.04$) in hemodialysis patients(8). The authors noted this was a reversal from what had previously been reported in medication adherence studies(14,15), and speculated sample size, follow-up period, or random phenomenon might apply. The other publication reported findings from a small 2014 Duquesne University study at an urban indigent primary care clinic, wherein medication discrepancies were more likely to persist in Caucasian subjects when compared to African Americans, despite pharmacist-led medication reconciliation. The authors theorized this finding might stem from variations in providers' communication styles with the two patient groups, but noted additional investigations in this area are needed(13).

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF;
  OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**

Affects large numbers, A leading cause of morbidity/mortality, Frequently performed procedure, High resource use, Patient/societal consequences of poor quality, Severity of illness

1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**

Medication management is a widely acknowledged problem in health care, generally(12-17), but is especially important for patients with ESRD, who often require 10 or more medications and take an average of 17-25 doses per day(1). Reducing MRPs has the potential to significantly reduce morbidity and mortality in dialysis-dependent patients. While there is a general paucity of pertinent randomized controlled trials (RCTs) in this area, one such study demonstrated an association between an increased focus on medication management practices and the identification of actual and potential MRPs, a decrease in the mean numbers of drugs taken by patients, and a reduction in all-cause hospitalization rates and hospital lengths-of-stay(1,9). Likewise, the IBPiD Study, a cross-sectional staff survey of three dialysis organizations comparing the perceived quality of patient-, provider-, and facility-level practices with SMR scores from USRDS facility reports, revealed units with lower-than-expected mortality rates convene multidisciplinary conferences sooner after dialysis patients return to the facility after hospitalization and perform medication reconciliation more frequently than high-mortality units(9). Finally, improved medication management practices will likely reduce healthcare costs. For example, a 2002 report estimated that every dollar spent on detecting and addressing MRPs in the dialysis population might ultimately save the healthcare system four dollars(10). More recently, a Minnesota study observed the reduction in total annual health expenditures exceeded the cost of providing MTM services by more than 12 to 1 in the general population. These savings would accrue from decreased prescription costs, from avoidance of unnecessary and/or inappropriate medications, and fewer hospitalizations(11).

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

1. Hakim RM, Collins AJ. Reducing avoidable rehospitalization in ESRD: A shared accountability. JASN. 2014;25(9):1891-1893.

2. Cardone KE, Bacchus S, Assimon MM, Pai AB, Manley HJ. Medication-related problems in CKD. Adv Chronic Kidney Dis. 2010;17(5):404-412.

3. Shoemaker SJ, Hassoi A. Understanding the landscape of MTM programs for Medicare Part D: Results from a study for the Centers for Medicare & Medicaid Services. J Am Pharm Assoc. 2011;51(4):520-526.

4. Forum of ESRD Networks' Medical Advisory Council. Medication Reconciliation Toolkit. 2009. Available at: http://esrdnetworks.org. Accessed March 22, 2016.

5. Parker WM and Cardone KE. Medication Management Services in a Dialysis Center: Patient and Dialysis Staff Perspectives. Albany College of Pharmacy and Health Services. January 2015. Available at: http://www.acphs.edu. Accessed March 22, 2016.

6. National Kidney and Urologic Diseases Information Clearinghouse. Kidney Disease Statistics for the United States. June 2012.

7. Pai AB, Cardone KE, Manley HJ, St. Peter WL, Shaffer R, Somers M, Mehrotra R. Dialysis Advisory Group of American Society of Nephrology. Medication reconciliation and therapy management in dialysis-dependent patients: Need for a

systematic approach. CJASN. 2013;8(11):1988-1999.

8. Manley HJ, Drayer DK, McClaran M, Bender W, Muther RS. Drug record discrepancies in an outpatient electronic medical record: Frequency, type, and potential impact on patient care at a hemodialysis center. Pharmacotherapy. 2003;23(2):231-239.

9. Pai AB, Boyd A, Depczynski J, Chavez IM, Khan N, Manley H. Reduced drug use and hospitalization rates in patients undergoing hemodialysis who received pharmaceutical care: A 2-year, randomized, controlled study. Pharmacotherapy. 2009; 29: 1433–1440.

10. Spiegel B, Bolus R, Desai AA, Zagar P, Parker T, Moran J, Solomon MD, Khawar O, Gitlin M, Talley J, Nissenson A. Dialysis practices that distinguish facilities with below- versus above-expected mortality. CJASN. 2010;5:2024-2033.

11. Manley HJ, Carroll CA. The clinical and economic impact of pharmaceutical care in end-stage renal disease patients. Semin Dial. 2002;15:45–49.

12. Isetts BJ, Schondelmeyer SW, Artz MB, Lenarz LA, Heaton AH, Wadd WB, Brown LM, Cipolle RJ. Clinical and economic outcomes of medication therapy management services: The Minnesota experience. J Am Pharm Assoc. 2008;48:203–211.

13. Stewart AL, Lynch KJ. Medication discrepancies despite pharmacist led medication reconciliation: The challenges of maintaining an accurate medication list in primary care. Pharm Pract. 2014;12(1)360.

14. Bedell SF, Jabbour S, Goldberg R et al. Discrepancies in the use of medications: Their extent and predictors in an outpatient practice. Arch Intern Med. 2000;160:2129-2134.

15. Wagner MM, Hogan WR. The accuracy of medication data in an outpatient electronic medical record. J Am Med Inform Assoc. 1996;3:61-68.

16. Cipolle RJ. Clinical and economic outcomes of medication therapy management services: The Minnesota experience. J Am Pharm Assoc. 2008;48:203–211.

17. Tache SV, Sonnichsen A, Ashcrof, DM. Prevalence of adverse drug events in ambulatory care: A systematic review. The Annals of Pharmacotherapy. 2001; 45(7-8):977-989.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
Not applicable.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria 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):*
Renal, Renal : End Stage Renal Disease (ESRD)

De.6. **Cross Cutting Areas** *(check all the areas that apply):*

Care Coordination, Safety, Safety : Medication Safety

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.)*
http://www.kidneycarepartners.com/files2/94

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)*
No data dictionary **Attachment:**

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Not applicable; new measure.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Number of patient-months for which medication reconciliation was performed and documented by an eligible professional during the reporting period.

The medication reconciliation MUST:
• Include the name or other unique identifier of the eligible professional;

AND

• Include the date of the reconciliation;

AND

• Address ALL known home medications (prescriptions, over-the-counters, herbals, vitamin/mineral/dietary (nutritional) supplements, and medical marijuana);

AND

• Address for EACH home medication: Medication name(1), indication(2), dosage(2), frequency(2), route of administration(2), start and end date (if applicable)(2), discontinuation date (if applicable)(2), reason medication was stopped or discontinued (if applicable)(2), and identification of individual who authorized stoppage or discontinuation of medication (if applicable)(2);

AND

• List any allergies, intolerances, or adverse drug events experienced by the patient.

_____

1. For patients in a clinical trial, it is acknowledged that it may be unknown as to whether the patient is receiving the therapeutic agent or a placebo.

2. “Unknown” is an acceptable response for this field.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
12 months.

S.6. **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, 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.
NUMERATOR STEP 1. For each patient meeting the denominator criteria in the given calculation month, identify all patients with each of the following three numerator criteria (a, b, and c) documented in the facility medical record to define the numerator for that month:

A. Facility attestation that during the calculation month:

1. The patient's most recent medication list in the dialysis medical record was reconciled to one or more external list(s) of medications obtained from the patient/caregiver (including patient-/caregiver-provided "brown-bag" information), pharmacotherapy information network (e.g., Surescripts®), hospital, or other provider AND that ALL known medications (prescriptions, OTCs, herbals, vitamin/mineral/dietary [nutritional] supplements, and medical marijuana) were reconciled;

AND

2. ALL of the following items were addressed for EACH identified medication:
   a) Medication name;
   b) Indication (or "unknown");
   c) Dosage (or "unknown");
   d)Frequency (or "unknown");
   e) Route of administration (or "unknown");
   f) Start date (or "unknown");
   g) End date, if applicable (or "unknown");
   h) Discontinuation date, if applicable (or "unknown");
   i) Reason medication was stopped or discontinued, if applicable (or "unknown"); and
   j) Identification of individual who authorized stoppage or discontinuation of medication, if applicable (or "unknown");

AND

3. Allergies, intolerances, and adverse drug events were addressed and documented.

B. Date of the medication reconciliation.

C. Identity of eligible professional performing the medication reconciliation.

NUMERATOR STEP 2. Repeat "Numerator Step 1" for each month of the one-year reporting period to define the final numerator (patient-months).

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Total number of patient-months for all patients permanently assigned to a dialysis facility during the reporting period.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, Populations at Risk : Individuals with multiple chronic conditions

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
DENOMINATOR STEP 1. Identify all in-center and home hemodialysis and peritoneal dialysis patients permanently assigned to the dialysis facility in the given calculation month.

DENOMINATOR STEP 2. For all patients included in the denominator in the given calculation month in "Denominator Step 1", identify and remove all in-center hemodialysis patients who received < 7 dialysis treatments in the calculation month.

DENOMINATOR STEP 3. Repeat "Denominator Step 1" and "Denominator Step 2" for each month of the one-year reporting period.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
In-center patients who receive < 7 hemodialysis treatments in the facility during the reporting month.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
As detailed in "Denominator Step 2" above, transient patients, defined as in-center patients who receive < 7 hemodialysis treatments in the facility during the reporting month, are excluded from the measure.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
Not applicable.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Not applicable.

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
Not applicable.

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Higher score

S.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)
Scores are calculated using the following algorithm. For each calculation month in the one-year reporting period:

1. IDENTIFY THE "RAW DENOMINATOR POPULATION"
Identify all in-center and home hemodialysis and peritoneal dialysis patients permanently assigned to the dialysis facility during the given calculation month.

2. REMOVE PATIENTS MEETING MEASURE EXCLUSION CRITERIA TO DEFINE THE "FINAL DENOMINATOR POPULATION" FOR THE CALCULATION MONTH
For all patients included in the denominator during the given calculation month in Step 1 above, identify and remove all in-center patients who received < 7 hemodialysis treatments during the given calculation month.

3. IDENTIFY THE "NUMERATOR POPULATION" FOR THE CALCULATION MONTH
For each patient remaining in the denominator during the given calculation month after Step 2, identify all patients with each of the following three numerator criteria (a, b, and c) documented in the facility medical record to define the numerator for that month:

A. Facility attestation that during the calculation month:

1. The patient's most recent medication list in the dialysis medical record was reconciled to one or more external list(s) of medications obtained from the patient/caregiver (including patient-/caregiver-provided "brown-bag" information), pharmacotherapy information network (e.g., Surescripts®), hospital, or other provider AND that ALL known medications (prescriptions, OTCs, herbals, vitamin/mineral/dietary [nutritional] supplements, and medical marijuana) were reconciled;

AND

2. ALL of the following items were addressed for EACH identified medication:
a) Medication name;
b) Indication (or "unknown");
c) Dosage (or "unknown");
d) Frequency (or "unknown");
e) Route of administration (or "unknown");
f) Start date (or "unknown");
g) End date, if applicable (or "unknown");
h) Discontinuation date, if applicable (or "unknown");
i) Reason medication was stopped or discontinued, if applicable (or "unknown"); and
j) Identification of individual who authorized stoppage or discontinuation of medication, if applicable (or "unknown");

AND

3. Allergies, intolerances, and adverse drug events were addressed and documented.

B. Date of medication reconciliation.

C. Identity of eligible professional performing medication reconciliation.

4. CALCULATE THE PERFORMANCE SCORE FOR THE CALCULATION MONTH
Calculate the facility's performance score for the given calculation month as follows:

Month's Performance Score = Month's Final Numerator Population ÷ Month's Final Denominator Population

5. CALCULATE THE ANNUAL PERFORMANCE SCORE
Calculate the facility's annual performance score as follows:

Facility's Annual Performance Score = (Facility's Month 1 Score + Month 2 Score +..... + Month 12 Score) ÷ 12

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Not applicable.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not applicable.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Medication Reconciliation for Patients Receiving Care at Dialysis Facilities is constructed as an "all or nothing" measure, such that a medication reconciliation event for which any of the numerator data elements are missing does not meet the measure criteria and is counted as a measure "fail" for that calculation month. Consequently, there is no missing data to report on this measure.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Dialysis facility medical record; intended for use by CMS in its CROWNWeb ESRD Clinical Data Repository.

S.25. **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.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Dialysis Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
Not applicable.

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
MM-2_NQF_TestingAttachment05-10-16FINAL.pdf

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2---2b7)

Measure Number: NQF 2988

Measure Title: ***Medication Reconciliation for Patients Receiving Care at Dialysis Facilities.***

Date of Submission: 5/10/2016 **Type of Measure:**

• Composite – ***STOP – use composite testing form***	• Outcome (*including PRO---PM*)
7. Cost/resource	☒ Process
12. Efficiency	□ Structure

Instructions

- **Measures must be tested for all the data sources and levels of analyses that are specified.** *If there is more than one set of data specifications or more than one level of analysis, contact NQF staff* **about how to present all the testing information in one form.**
- For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.
- For outcome and resource use measures**, section** 2b4 **also must be completed.**
- **If specified for** multiple data sources/sets of specificaitons **(e.g., claims and EHRs), section** 2b6 **also must be completed.**
- **Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2---2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.**
- **If you are unable to check a box, please highlight or shade the box for your response.**
- **Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins).**

Contact NQF staff if more pages are needed.

- **Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.**
- **For information on the most updated guidance on how to address sociodemographic variables and testing**

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] **demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For** PRO---PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.**

2b2. Validity testing [11] **demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For** PRO---PMs and composite performance measures**, validity should be demonstrated for the computed performance score.**

2b3. **Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion;** [12]
AND
If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category

computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence---based risk---adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful differences in performance**; [16]

OR

there is evidence of overall less---than---optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO---PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter---rater/abstractor or intra---rater/abstractor studies; internal consistency for multi---item scales; test---retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal---to---noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less---than---optimal performance may not demonstrate much variability across providers.

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.23*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☒ clinical database/registry	☒ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☒ other: **CROWNWeb ESRD Clinical Data Repository**	☐ other: Click here to describe

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

The *M*edication *Reconciliation for Patients Receiving Care at Dialysis Facilities* measure was tested using data from three KCQA member dialysis organizations, each with the capacity to provide retrospective analyses from a data warehouse/repository. All pertinent data from all eligible patients (i.e., adult and pediatric in---center and home hemodialysis and peritoneal dialysis) of the participating organizations during the testing period were included in the datasets.

1.3. What are the dates of the data used in testing? **April 1---September 30, 2015.**

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.26*)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☒ other: **Dialysis facility**	☒ other: **Dialysis facility**

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

The measured entity is the dialysis facility. All facilities in each of the three participating dialysis organizations were included in the analysis. The number of contributing facilities varied by month, but was approximately 5,292 facilities in each of the six months of the study. The range of contributing facilities was 5,258 (April 2015) to 5,319 (September 2015).

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*):

All patients (i.e., adult and pediatric in---center and home hemodialysis and peritoneal dialysis) in all facilities in each of the three participating dialysis organizations were included in the analysis. This translated to approximately 323,000 to 328,000 patients for each of the six months of the study period. Demographic information such as age, sex, and race were not assessed.

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.

Not applicable.

1.8 What were the patient---level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient---reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Sociodemographic information such as income, education, and language were not assessed.

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*):

Empirical reliability testing at the measure score level was conducted using the beta---binomial test (1) on the data pulls from the three participating dialysis organizations. Each organization pulled Q2 and Q3 data for 2015 for all facilities in accordance with the measure specifications, then provided their datasets for each facility (anonymized) for each month to an independent methodologist.

The beta---binomial method is characterized as a "natural model for estimating the reliability of simple pass/fail rate measures," and so is appropriate for this KCQA metric. Using this approach, reliability represents the ability of a measure to effectively distinguish the performance of one measured entity from another. The model is based on the beta distribution for the "true" scores for the measured entity, and assumes the entity's score is a binomial random variable conditional on the entity's true value that comes from the beta distribution. The beta distribution, which can be symmetric, skewed, or U---shaped, is "a very flexible distribution on the interval from 0 to 1" (1).

Reliability as calculated for the KCQA measure is thus the ratio of signal to noise, where the signal is the proportion of the variability in measured performance that can be explained by real differences in performance. A reliability of 0 implies that all variability in a measure is attributable to measurement error, while a reliability of 1 implies all the variability is attributable to real differences in performance. The higher the reliability score, the greater the confidence the measure distinguishes the performance of one dialysis facility from another. A reliability statistic of 0.7 is generally viewed as an acceptable threshold (1).

1. **Adams, JL. The reliability of provider profiling: A tutorial. RAND Health, 2009.**

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*):

For the 6---month study period, for all facilities (excluding those with <=11 patients in a given reporting month, as per the measure specifications [approximately 3.7% of facilities each month]), the mean reliability of the measure is 0.9935 (range = 0.8166---1). (Results also contained in the KCQA Testing Data Attachment.)

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?*)

As previously noted, a reliability statistic of 0.7 is generally viewed as an acceptable threshold (1). Our reliability statistic of 0.9935 is excellent, suggesting the measure is highly reliable and effectively differentiates real differences in performance among facilities.

2b2. VALIDITY TESTING

2b2.1. 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*)

2b2.2. 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):*

Per NQF guidance (2), face validity of the measure was assessed through a systematic and transparent process by identified experts. Specifically, two separate groups of experts in the field of ESRD and dialysis care were identified—lead (voting) representatives from KCQA member organizations and a 9---member expert panel identified by the KCQA Steering Committee. Each group completed a face validity assessment that explicitly addressed whether performance scores resulting from the measure, as specified, provide an accurate reflection of quality. Individuals responded to the following two questions:

- **How likely is it that the measure score provides an accurate reflection of medication reconciliation quality? (highly unlikely; unlikely; neither likely nor unlikely; likely; highly likely)**

- **What is the likelihood that the measure can be used to distinguish good from poor quality? (highly unlikely; unlikely; neither likely nor unlikely; likely; highly likely)**

2. NQF. *Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement.* April 2015. Available at: http://www.qualityforum.org/Projects/i---m/Measure_Evaluation_Guidance/Measure_Evaluation_Guidance.aspx. Accessed March 22, 2016.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t---test*):
The face validity assessment yielded the following:

- **KCQA Member Organizations' Lead Representatives:**
  1. **77.3% of KCQA Lead Representatives (n=22) agreed it is highly likely or likely that the measure score provides an accurate reflection of medication reconciliation quality.**
  2. **77.3% of the panel agreed that it is likely/highly likely that the measure can be used to distinguish good from poor quality.**

- **Expert Panel:**
  1. **88.9% of the 9---member panel agreed it is highly likely or likely that the measure score provides an accurate reflection of medication reconciliation quality.**
  2. **77.8% of the panel agreed it is highly likely or likely that the measure can be used to distinguish good from poor quality.**

(Results also contained in the KCQA Testing Data Attachment.)

2b2.4. 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?*):
Both the Expert Panel and KCQA Lead Representatives showed significant agreement that scores from the measure as specified will accurately reflect medication reconciliation quality and will differentiate quality among providers. Our interpretation of these results is that this measure has substantial face validity.

2b3. EXCLUSIONS ANALYSIS
NA ☐ no exclusions

2b3.1. 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*):
Medication Reconciliation for Patients Receiving Care at Dialysis Facilities **has one patient---level and one facility---level exclusion:**

- **Patient---Level Exclusion: Transient patients, i.e., in---center hemodialysis patients who receive <7 dialysis treatments in the facility during the calculation month.**
- **Facility---Level Exclusion: Facilities with <=11 (i.e., <12) patients during the calculation month.**

For the patient---level exclusion, the analysis was conducted on the data pulls from the three participating dialysis organizations. Again, each participating organization pulled 2015 Q2 and Q3 data for all facilities in accordance with the measure specifications, then provided their datasets for each facility (anonymized) for each month. For each facility across the three participating dialysis organizations, the overall number and percentages of patients meeting the exclusion (transient patients, i.e., in---center hemodialysis patients who receive <7 treatments during the calculation

month) was recorded for each of the 6 months. The combined dataset was then examined to identify the monthly and overall frequencies of the occurrence, as well as the variability of the exclusion.

The facility---level exclusion parameter of <=11 was empirically determined during testing specifically to assess the impact on reliability of a "small numbers" effect. The effect of the measure's reliability in the context of excluding facilities at varying thresholds, including CMS's general implementation approach of excluding facilities with <11 patients/patient events (I.e., <=10). Both the percentage of facilities that would be excluded from measurement, as well as the reliability of the measure for small facilities were analyzed:

- **Using the CMS <11 threshold resulted in the exclusion of 3.3---3.6% of facilities from the measure, depending on the month.**
- **<11 threshold reliability statistics: Minimum = 0.3615; 10th Percentile = 0.6937; Median = 0.9174; 90th Percentile = 1; Maximum = 1**
- **At the 10th percentile, the measure does not achieve the previously cited reliability threshold of 0.7.**

Additional analyses were performed to determine the sample size that would yield a reliability statistic of 0.7 for all but outliers (defined as below the 10th percentile). Based on these analyses, a reliability statistic of at least 0.7 for the 10th percentile occurs at the threshold of <=11 (i.e., <12) patients in a given reporting month:

- **<=11 threshold reliability statistics: Minimum = 0.3622; 10th Percentile = 0.7089; Median = 0.9177; 90th Percentile = 1; Maximum = 1**

Based on this analysis, KCQA specified "facilities with <=11 (i.e., <12) patients in the reporting month" as the empirically appropriate small---numbers exclusion for the measure.

2b3.2. 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*):

Findings for the exclusions analysis are as follows:

- **Transient in---center hemodialysis patients who receive <7 dialysis treatments in the facility during the reporting month:**
  - **Mean number of patients excluded per facility in each study month: April = 1.95; May = 1.94; June = 1.96; July = 1.93; Aug = 1.94; Sep = 1.93**
  - **Mean number of patients excluded per facility, per month = 1.94**
  - **Total number (and percent) of patients excluded across all facilities in each study month: April = 8,972 (2.79%); May = 8,949 (2.77%); June = 9,073 (2.80%); July = 8,942 (2.74%); Aug = 9,007 (2.76%); Sep = 8,986 (2.75%)**
  - **Mean number (and percent) of patients excluded across all facilities, per month: 8,988 (2.77%)**
  - **Total number (and percent) of patient---months excluded across all facilities over the 6---month study period = 53,928 (2.77%)**

- **Facilities with <=11 patients during the reporting month:**
  - **Number (and percent) of facilities excluded in each study month: April = 180 (3.76%); May = 185 (3.86%); June = 177 (3.68%); July = 181 (3.76%); Aug = 179 (3.71%); Sep = 180 (3.72%)**

- **Mean number (and percent) of facilities excluded over the 6---month study period = 180.33/3.75%**

2b3.3. 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 frequency and variability with which the exclusions were encountered during testing is sufficient to demonstrate they are necessary to prevent unfair distortion of performance results.

2b4. 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 2b5.

2b4.1. What method of controlling for differences in case mix is used?
☒ **No risk adjustment or stratification**
- ☐ **Statistical risk model with** Click here to enter number of factors **risk factors**
- ☐ **Stratification by** Click here to enter number of categories **risk categories**
- ☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use 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.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

2b4.4a. What were the statistical results of the analyses used to select risk factors?

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

2b4.5. 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 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c---statistic, R---squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer---Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves: 2b4.9.

Results of Risk Stratification Analysis:

2b4.10. 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*)

2b4.11. 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*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. 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):*

Descriptive statistics for the performance measure scores for all tested entities (facilities) were constructed. These statistics include the mean, standard deviation and standard error, 95% confidence interval, median, mode, range of scores, and the interquartile range of scores across the measured entities.

Meaningful difference is defined as a significant spread (>20%) between minimum and maximum scores or a significant spread between median and minimum scores, median and maximum scores, and/or the interquartile range.

2b5.2. 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*):

Descriptive statistics for the performance measure scores are as follows:

- **Mean Performance Score = 52.62%**
- **Standard Deviation = 32.83**
- **Standard Error = 0.197**
- **95% Confidence Interval = 52.24 to 53.01**
- **Median Score = 48.18**
- **Mode of Scores = 100**
- **Range of Scores = 0 to 100**
- **Interquartile Range = 27.59 to 87.62**

2b5.3. 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?*)

Results are interpreted as showing a significant spread between the minimum and maximum scores (0---100), as well as the median and minimum (0---48.14) and maximum scores (48.14---100) and the interquartile range, indicating that the measure identifies clinically and practically meaningful

differences in performance among the measured entities.

2b6. 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 SDS 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 SDS 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.***

2b6.1. 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*)

2b6.2. 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*)

2b6.3. 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*)

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*):
Medication Reconciliation for Patients Receiving Care at Dialysis Facilities **is constructed as an "all or nothing" measure, such that an event for which any of the numerator data elements are missing does not meet the measure criteria and is counted as a measure "fail" for that patient for that month. Consequently, there are no missing data to report on this measure.**

2b7.2. 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*):
Not applicable, as noted above.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i*.e., what do the results mean in terms*

of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data):
Not applicable, as noted above.

3. Feasibility

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

3a. Byproduct of Care Processes

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

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score)
If other:

3b. Electronic Sources

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

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*
ALL data elements are in defined fields in electronic health records (EHRs)

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.**

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.**
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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
MEDICATION MANAGEMENT DEFINITIONAL DISCREPANCIES. When developing the measure specifications and operationalizing the specifications for testing, it was noted while all three dialysis organizations that participated in testing have identified and engage in the same three components of medication management—i.e., documentation, reconciliation, and review—one organization defined reconciliation and review in reverse to those detailed in the KCQA measure specifications. Specifically, "medication reconciliation" is defined within that organization as "the process of creating the most accurate list of all medications that the patient is taking by comparing the most recent medication list in the medical record to one or more external list(s) of medications obtained from a patient or caregiver," while "medication review" is defined as "a process of evaluating a patient's medications and confirming them as being appropriate, safe, and convenient for the patient; a review with the patient may be included."
Based on other KCQA Workgroup member input and our outreach to the other two testing organizations and KCQA members, however, this appeared to be an outlier situation—albeit a significant one. Our final approach to the medication management definitions was ultimately agreed upon because the majority of dialysis organizations use this convention, as do hospitals, pharmacists, and the existing NQF-endorsed measures in the area.

DATA SYSTEM DISCREPANCIES. Again, when developing the measure specifications and operationalizing the specifications for testing, variations between the electronic medical record systems of the three large dialysis organizations that participated in testing were identified. For instance, a given data element (e.g., indication, start date, name of eligible professional) might not be present or might be available only as a free text field. It was further noted that this variability might be even greater in the medium and small dialysis organizations. Given the variability among electronic systems and because some medications are prescribed by other entities for which "indication" may be unknown, for example, it was determined that "unknown" must be an allowable response to many data elements so as to maintain the measure's feasibility.

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)*.
Not applicable.

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.

Planned	Current Use (for current use provide URL)
Public Reporting Payment Program Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Quality Improvement (Internal to the specific organization) Not in use	

4a.1. **For each CURRENT use, checked above, provide:**

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

Not applicable; this is a new measure that is not yet in use as specified. Variants of the measure are currently in use by KCQA member dialysis organizations for internal quality improvement, prompting KCQA to develop this measure to standardize the specifications and definitions for accountability purposes.

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
This is a new measure undergoing initial endorsement assessment. The measure is not yet in use as specified; however, variants of the measure are currently in use by KCQA member dialysis organizations for internal quality improvement, prompting KCQA to

develop this measure to standardize the specifications and definitions for accountability purposes.

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

The measure was developed for use by CMS for its accountability initiatives. We note the measure requires a number of data fields not currently available in the CROWNWeb ESRD clinical data repository, and would require a system update for implementation. As we have done for other KCQA measures, we intend to commence discussions with CMS in this regard, specifically to request that the measure be included in the Measures Under Consideration for Use in Federal Programs List submitted to NQF's Measure Applications Partnership (MAP) in an upcoming cycle and that a CROWNWeb System Change form be created to commence building the necessary data elements into the system.

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

Not applicable; new measure undergoing initial endorsement review.

4b.2. **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 new and is not yet in use. However, variants of the measure are currently in use by KCQA member dialysis organizations for internal quality improvement. Standardizing specifications and definitions for accountability purposes will improve and expedite identification and resolution of real and potential medication-related problems (MRPs) in ESRD patients. Associated hospitalization, readmissions, mortality, and health care costs should consequently be minimized.

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**

No unintended consequences were identified during testing.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

Yes

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

0097 : Medication Reconciliation Post-Discharge
0554 : Medication Reconciliation Post-Discharge (MRP)
2456 : Medication Reconciliation: Number of Unintentional Medication Discrepancies per Patient

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

5a. Harmonization

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 completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
Medication Reconciliation for Patients Receiving Care at Dialysis Facilities is harmonized with existing NQF-endorsed medication reconciliation measures in that all similarly specify that the medication reconciliation must address ALL prescriptions, over-the-counters,herbals,vitamin/mineral/dietary (nutritional) supplements AND must contain the medications' name, dosage, frequency, and route. The KCQA measure, however, is unique among the currently endorsed medication reconciliation measures in that the level of analysis is the dialysis facility. The KCQA measure also moves beyond a single "check/box", specifying multiple components that must be met to be counted as a "success." It requires the following additional information on each medication, where applicable and known: indication, start and end date, discontinuation date, reason the medication was stopped or discontinued, and identification of the individual who authorized stoppage or discontinuation of the medication. Additionally, given the increasing frequency with which medical marijuana is prescribed, the KCQA measure specifies that this pharmacotherapeutic agent must be addressed during the reconciliation. KCQA believes these additional foci are necessary to ensure the medication reconciliation process is as comprehensive as possible to better identify and effectively address potential sources of adverse drug-related events and not function merely as a single "check-box" measure. Testing demonstrated these data elements are effectively captured and recorded in facility's electronic medical record systems during the routine medication reconciliation process.

5b. Competing Measures

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

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
Not applicable; this medication management measure is unique in its specific focus on the ESRD population.

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** tbKCQA_Specs-TestingData05-10-16FINAL.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Kidney Care Quality Alliance (KCQA)
Co.2 **Point of Contact:** Lisa, McGonigal, lmcgon@msn.com, 203-530-9524-
Co.3 **Measure Developer if different from Measure Steward:** Kidney Care Quality Alliance (KCQA)
Co.4 **Point of Contact:** Lisa, McGonigal, lmcgon@msn.com, 203-530-9524-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.

The KCQA Steering Committee guides the measure development process. Steering Committee members are:

- Edward Jones, MD; KCQA Co-Chair — Renal Physicians Association
- Allen Nissenson, MD; KCQA Co-Chair — DaVita
- Jason Spangler, MD, MPH — Amgen
- Donna Bednarski, RN, MSN — American Nephrology Nurses Association
- Barbara Fivush, MD — American Society of Pediatric Nephrology
- Raymond Hakim, MD, PhD — American Society of Nephrology
- Scott Ash, MHA — Fresenius Medical Care North America
- Chris Lovell, RN, MSN — Dialysis Clinics, Inc.
- Thomas Manley, RN, BSN — National Kidney Foundation
- Gail Wick, MHSA, BSN, RN — American Kidney Fund
- Shari M. Ling, MD, Chief Medical Officer, Centers for Medicare and Medicaid Services, Center for Clinical Standards and Quality (CCSQ) – CMS Liaison Member

The KCQA Measure Feasibility/Testing Workgroup provided technical expertise and guidance to develop the specifications. Workgroup members were:

- Richard Faris, PhD, MSc, RPh — DaVita
- James Guffey — Dialysis Patient Citizens
- Jeffrey Hymes, MD — Fresenius Medical Care North America
- Len Usvyat, PhD — Fresenius Medical Care Renal Therapies Group
- Harold Manley, PharmD, FASN, FCCP — Dialysis Clinics, Inc.
- Paul Miller, MD — Renal Physicians Association
- Donald Molony, MD — Forum of ESRD Networks
- Glenda Payne, MS, RN, CNN — American Nephrology Nurses Association
- Sharon Perlman, MD — American Society of Pediatric Nephrology
- Wendy St. Peter, PharmD, FASN, FCCP, FNKF — National Kidney Foundation
- Gail Wick, MHSA, BSN, RN; KCQA Steering Committee Liaison — American Kidney Fund

KCQA Lead (Voting) Representatives identify KCQA's measure development foci, review the Workgroup's output and testing results, and approve major milestones during the development of the process, including and assessment of the face validity of the measure and submission to NQF. KCQA Lead Representatives are:

- Michael Heiffets, MD — AbbVie
- Qing Zuraw, MD, MBA — Akebia Therapeutics, Inc.
- Gail Wick, MHSA, BSN, RN — American Kidney Fund
- Glenda Payne, MS, RN, CNN — American Nephrology Nurses' Association
- Richard Cronin, MD — American Renal Associates, Inc.
- Raymond Hakim, MD, PhD — American Society of Nephrology
- Barbara Fivush, MD — American Society of Pediatric Nephrology
- Jason Spangler, MD, MPH — Amgen
- Maggie Gellens — Baxter Healthcare Corporation
- RJ Picciano — Board of Nephrology Examiners and Technology
- Peter DeOreo, MD — Centers for Dialysis Care
- LeAnne Zumwalt — DaVita Healthcare Partners, Inc.
- James Michael Guffey — Dialysis Patient Citizens
- Doug Johnson, MD — Dialysis Clinic, Inc.
- Jeffrey Hymes, MD — Fresenius Medical Care North America
- Robert Kossman, MD — Fresenius Medical Care Renal Therapies Group
- Jennifer Holcomb/William Poire — Greenfield Health Systems
- Thomas Nusbickel — Hospira
- Greg Madison — Keryx Biopharmaceuticals, Inc.
- Cherilyn Cepriano — Kidney Care Council
- Linda Keegan — Kidney Care Partners
- Donald Molony, MD/Andrew Howard, MD — The National Forum of ESRD Networks
- Tonya Saffer — National Kidney Foundation
- Deb Cote — National Renal Administrators Association
- Nancy Gallagher — Nephrology Nursing Certification Commission
- Tosha Whitley — Northwest Kidney Centers

- Leslie Spry, MD — NxStage Medical
- Paul Palevsky, MD — Renal Physicians Association
- Jonathan Lorch, MD — Rogosin Institute
- Sara Froelich — Sanofi
- Brigitte Schiller — Satellite Healthcare
- Stan Lindenfeld, MD — U.S. Renal Care

In addition to the assessment by KCQA Lead Representatives, KCQA conducted face validity assessment at the performance score level by convening a 9-member panel of other renal experts:

- Lorien Dalrymple, MD, MPH — University of California, Davis Health System
- Norma Gomez, MSN, MBA — Satellite Healthcare
- Hrant Jamgochian, JD, LLM — Dialysis Patient Citizens
- Charla Litton, FNP — People's Health Network
- Klemens Meyer, MD — Dialysis Clinic, Inc.
- Donna Painter, RN — Fresenius Medical Care North America
- Barry Smith, MD — Rogosin Institute
- Katherine Swanzy — DaVita Kidney Care
- Daniel Weiner, MD, MS — Tufts Medical Center

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2016

Ad.3 **Month and Year of most recent revision:** 05, 2016

Ad.4 **What is your frequency for review/update of this measure?** Annually, and as needed with changes or additions to the evidence base.

Ad.5 **When is the next scheduled review/update for this measure?** 05, 2017

Ad.7 **Disclaimers:** Dialysis facility performance measures (Measures) and related data specifications, developed by the Kidney Care Quality Alliance (KCQA), primarily funded by Kidney Care Partners, are intended to facilitate quality improvement activities by dialysis providers.

These Measures are intended to assist dialysis facilities in enhancing quality of care. Measures are designed for use by any dialysis facility. These performance Measures are not clinical guidelines and do not establish a standard of medical care. KCQA has not tested its Measures for all potential applications. KCQA encourages the evaluation of its Measures.

Measures are subject to review and may be revised or rescinded at any time by KCQA. The Measures may not be altered without the prior written approval of KCQA. Measures developed by KCQA, while copyrighted, can be reproduced and distributed, without modification, for noncommercial purposes, e.g., use by dialysis providers in connection with their care delivery or for research. Commercial use is defined as the sale, license, or distribution of the Measures for commercial gain, or incorporation of the Measures into a product or service that is sold, licensed or distributed for commercial gain. Commercial uses of the Measures require a license agreement between the user and Kidney Care Partners, on behalf of KCQA.

Neither KCQA nor its members shall be responsible for any use of these Measures.

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

Ad.8 **Additional Information/Comments:**

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 2993
Measure Title: Potentially Harmful Drug-Disease Interactions in the Elderly
Measure Steward: National Committee for Quality Assurance
Brief Description of Measure: The percentage of patients 65 years of age and older who have evidence of an underlying disease, condition or health concern and who are dispensed an ambulatory prescription for a potentially harmful medication, concurrent with or after the diagnosis. Four rates are reported for this measure:

-Rate 1: The percentage of those with a history of falls that received a potentially harmful medication
-Rate 2: The percentage of those with dementia that received a potentially harmful medication
-Rate 3: The percentage of those with chronic kidney disease that received a potentially harmful medication
-Rate 4: Total rate

A lower rate represents better performance for all rates.
Developer Rationale: Lowering the rate of potentially harmful drug-disease interactions in the elderly population should decrease morbidity and mortality associated with adverse drug reactions.

Numerator Statement: Numerator 1: Patients with a history of falls who received at least one potentially harmful medication from Table DDE-A or Table DDE-B
Numerator 2: Patients with a diagnosis of dementia who received at least one potentially harmful medication from Table DDE-D
Numerator 3: Patients with chronic kidney disease who received at least one potentially harmful medication from Table DDE-E
Numerator 4: The sum of the three numerators
Denominator Statement: All patients ages 65 years of age and older with a history of falls, dementia or chronic kidney disease in the measurement year or the year prior to the measurement year.
Denominator Exclusions: The following are exclusions for the condition-specific rates and total rate:

For those who meet denominator criteria for the history of falls rate (Rate 1): exclude those with a diagnosis of psychosis, schizophrenia, bipolar disorder or seizure disorder.

For those who meet denominator criteria for those with dementia rate (Rate 2): exclude those with a diagnosis of psychosis, schizophrenia or bipolar disorder.

Measure Type: Process
Data Source: Administrative claims, Electronic Clinical Data, Electronic Clinical Data : Pharmacy
Level of Analysis: Health Plan, Integrated Delivery System

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *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.

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 provides evidence based on the AGS Beers Criteria recommendations against the use of potentially harmful medications in older adults with specific conditions. The AGS Beers Criteria identifies 12 conditions where there are potentially inappropriate medications. This measure includes three of those conditions: history of falls or fracture, dementia or cognitive impairment, and chronic kidney disease. The time period covered by the body of evidence was 2004-14.

Guidance from the Evidence Algorithm

1-No→3-Yes→ 4-Yes→ 5a→HIGH

Questions for the Committee:

- *For process 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?*

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

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

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

- The developer provided data extracted from HEDIS data collection for Medicare Advantage Health Plans (including both HMO and PPO plans). The performance data is summarized at the health plan level. The data demonstrates variation in all four rates of the measure.
- For 2014, 48.0 percent of individuals with a history of falls received at least one high-risk medication. Among individuals with dementia, 48.5 percent received at least one high-risk medication and among those with chronic kidney disease, 9.6 percent received at least one high-risk medication.
- The national mean performance for the total rate was 41.5 percent. Overall, rates from 2013 to 2014 showed a slight decrease, yet the 2014 rates still suggest significant room for improvement, particularly for the history of falls and dementia rates.
- For all rates there is a sizeable gap between the plans at the 10th percentile and 90th percentile, demonstrating a gap in care between the best and worst performing health plans.

Disparities

- The developers did not stratify the measure by race, ethnicity, or language. They cited a study that demonstrated the difficulty of collecting valid data that would allow the assessment of health disparities at the health plan level. They did note that the measure can be stratified by demographic variables and the HEDIS Health Plan Measure Set contains two measures that can assist with stratification to assess healthcare

disparities.

- While disparities for this measure have not been well studied, there is some evidence to suggest that women are more likely to receive a potentially inappropriate medication than men. A retrospective cohort study of 966,000 men and women treated by the Veteran's Health Administration showed that women were more likely than men to receive medications that may have harmful interactions with chronic conditions as described by the Beers Criteria (Bierman et al., 2007). In a different study, a retrospective database analysis of HEDIS data from the Department of Veterans Affairs found that Hispanics and those with no copayments had higher rates of medications listed as potentially harmful than whites or those with required copayments (Pugh, 2011).

Questions for the Committee:

- *Specific question on information provided for gap in care.*
- *Is there a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*

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

Comments: ** This is a process measure that includes 4 separate rates. The developer provided a systematic review completed by the American Geriatrics Society 2015 Beers Criteria Update Expert Panel. The evidence for each element of the measure was graded as moderate or high based on AGS Beers Criteria.

** The evidence provides a high certainty that the elements measured in the measure prove greater risk than benefits to the targeted population for the measure. The process of taking high-risk medications in the elderly population has been demonstrated to correlate to higher outcome incidences of Falls, progression of dementia and kidney insult.

** This is a process measure. Lowering the rate of potentially harmful medications in the elderly who have specific conditions (history of falls/fracture, chronic kidney disease, dementia/cognitive impairment) should improve morbidity and mortality associated with adverse drug reactions. The evidence was reviewed and demonstrated quality, consistency and is of sufficient quantity. The Beer Criteria supports the developers recommendations against the use of potentially harmful medications in older adults with the specified conditions.

1b. Performance Gap

Comments: ** The measure has been in use by NCQA since 2007. However, performance data was only provided for 2013 and 2014. There are very slight improvements in each of the four rates. However, the developer noted that opportunities for improvement were most significant for individuals with a history of falls and individuals with dementia where almost half of patients received at least one high-risk medication contra-indicated based on diagnosis. Opportunity was least significant for individuals with chronic kidney disease where fewer than 10% received at least one high-risk medication.

** Yes, a high correlation exists between those taking a high risk medication and the outcomes measured. While there are many factors involved in the outcomes measured, high risk medications are a preventable or risk mitigated factor than can be influenced. Given the incidence of events, downstream clinical and economic effects and preventable nature of the measured constructs there is considerable evidence for a gap in care that can be influenced.

**The data provided did not provide sufficient evidence to identify a disparities issue.

**There is a significant performance difference between health plans at the 10th and 90th percentiles.

**The developer indicated that the measure is not stratified by race, ethnicity or language and that one study showed the difficulty of collecting data that would allow for an assessment of health disparities at the health plan level. Stratification by demographic variables could be done, and the HEDIS health plan measure set could assist with this process. One VA study showed that women are more likely to received medications with harmful interactions than men.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Administrative Claims

Specifications:

- This measure uses administrative claims data, including ICD-9, ICD-10, CPT, HCPCS, and UB revenue codes, to identify eligible patients who were prescribed a "potentially harmful medication."
- Data for the measure are collected via NCQA's online data submission system.
- The measure comprises four rates, each with its own numerator and denominator; the fourth rate is a sum of the previous three rates.
    - Rate 1 identifies patients with an accidental fall or hip fracture who were dispensed an ambulatory prescription for an antipsychotic, benzodiazepine, nonbenzodiazepine hypnotic or tricyclic antidepressant, or H2 receptor antagonist or anticholinergic agent.
    - Rate 2 identifies patients who were diagnosed with dementia or prescribed a dementia medication who were dispensed an ambulatory prescription for an antipsychotic, benzodiazepine, nonbenzodiazepine hypnotic or tricyclic antidepressant, or H2 receptor antagonist or anticholinergic agent.
    - Rate 3 identifies patients with a diagnosis of ESRD, stage 4 chronic kidney disease, or a kidney transplant who were dispensed an ambulatory prescription for a NSAID or Cox-2 selective NSAID.
    - The numerator for Rate 4 is the sum of the numerators for Rates 1, 2, and 3; the denominator for Rate 4 is the sum of the denominators for Rates 1, 2, and 3.
- The developer notes that patients with more than one disease or condition may appear in the measure multiple times (i.e., in each indicator for which they qualify).
- The developer has provided code sets for the data elements required to calculate the measure.

Questions for the Committee :

- *Specific questions on the specifications, codes, definitions, etc.*
- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

SUMMARY OF TESTING

Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing Beta-binomial testing

Results of reliability testing

Using 2014 Health Plan performance data, reliability for the rates in this measure are as shown below. Strong reliability is demonstrated since majority of variances is due to signal and not to noise

Rate	Beta Binomial Rate
Rate 1 (History of Falls)	0.96565
Rate 2 (Dementia)	0.97552
Rate 3 (Chronic Kidney Disease)	0.95273
Rate 4 (Total)	0.98571

Guidance from the Reliability Algorithm Based on these results, the reliability of this measure should be rated as "High".

Questions for the Committee:

- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

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

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

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, summarize the validity testing from the prior review:

This is described below.

SUMMARY OF TESTING

Validity testing level ☐ **Measure score** ☐ **Data element testing against a gold standard** ☒ **Both**

Method of validity testing of the measure score:

- ☐ **Face validity only**
- ☒ **Empirical validity testing of the measure score**

Validity testing method: There was both an assessment of face validity and also of construct validity by correlations of this measure with other measures of medication safety.

Validity testing results:

Results of Face Validity Assessment:

Step 1: This measure was developed to address potentially harmful drug-disease interactions in the elderly. NCQA and the GMAP worked together to assess conditions and medications based on the AGS Beers Criteria.

Step 2: The measure was field-tested from 2004-2005. After reviewing field test results the CPM recommended to send the measure to public comment with a majority vote in 2006.

Step 3: The measure was released for Public Comment in 2006 prior to publication in HEDIS. The CPM recommended moving this measure to first year data collection by a majority vote.

Step 4: The measure was introduced in HEDIS 2007. Organizations reported the measures in the first year and the results were analyzed for public reporting in the following year. The CPM recommended moving this measure public reporting with a majority vote.

Step 5: The measure is currently undergoing re-evaluation.

Conclusion: The measure was deemed to have the desirable attributes of a HEDIS measure in 2006 (relevance, scientific soundness, and feasibility).

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

The measure developer provided a Table in 2b3.2 describing the rate and distribution of exclusions across health plans reporting 2014 HEDIS data.

	Number of plans	Average rate of exclusions	Standard deviation	Min	25th	50th	75th	Max
Rate 1: History of Falls[1]	384	15.3	11.5	0.0	0.0	16.8	20.9	56.0
Rate 2: Dementia[2]	382	18.9	12.8	0.0	9.1	20.6	28.4	50.9
Rate 4: Total	409	14.8	10.7	0.0	6.9	16.0	20.67	61.0

[1]For the History of Falls rate, those with a diagnosis of psychosis, schizophrenia, bipolar disorder or seizure disorder are excluded.
[2]For the Dementia rate, those with a diagnosis of psychosis, schizophrenia or bipolar disorder are excluded.

Questions for the Committee:

- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

Variation in Performance across Plans (HEDIS Results using 2014 Data)

Rate	Number of Plans	Mean	SD	10th (Better)	25th	50th	75th	90th (Worse)	IQR	P-Value
Rate 1 (History of Falls)	387	48.0%	8.3	38.8%	43.1%	47.7%	51.9%	58.5%	8.8	0.0026
Rate 2 (Dementia)	385	48.5%	9.1	39.2%	42.8%	46.8%	52.8%	61.0%	10	0

Rate 3 (Chronic Kidney Disease)	356	9.6%	6.1	3.9%	5.8%	8.1%	12.0%	17.1%	6.2	0
Rate 4 (Total)	412	41.5%	7.9	33.5%	36.7%	40.3%	44.9%	51.3%	8.2	<0.001

IQR: Interquartile range
p-value: P-value of independent samples t-test comparing plans at the 25th percentile to plans at the 75th percentile.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

N/A

2b7. Missing Data

This was reported by the developer: Plans collect this measure using all administrative data sources. NCQA's audit process checks that plans' measure calculations are not biased due to missing data.

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

Committee pre-evaluation comments

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

2a1. & 2b1. Specifications

Comments: ** The measure utilizes administrative data to calculate each of the four rates and is very clearly specified. However, given that the measure is currently undergoing re-evaluation an ad hoc review may be necessary if significant changes are made to the measure specifications that would impact the reliability and/or validity.
** The data elements are clearly defined with appropriate coding provided. The logic algorithm is clear and this measure should be able to implemented consistently.
** The data source is administrative claims; specifications are clear. Specifications are consistent with the evidence

2a2. Reliability Testing
Comments: ** Reliability testing was performed on each of the four measure scores. Each of the four rates had a reliability score close to 1 - ranging from 0.95273 for chronic kidney disease to 0.98571 for the total rate - indicating that the measure is highly reliable.
** Administrative claims data were utilized from over 400 health plans from the HEDIS 2014 data set. The health plans were geographically diverse and varied in size. Reliability across the four indicators within the measure have a reliability above 0.95.
** Beta-binomial testing demonstrated strong reliability since the majority of variances is due to signal and not to noise. The testing level was the measure score and testing was performed with the data source and level of analysis indicated. High reliability was demonstrated.

2b2. Validity Testing
Comments: ** Both face and construct validity testing were performed.
**The developer utilized its internal HEDIS measure life cycle process to determine face validity. The measure went through a series of expert panel review and public comments. The measure passed through the HEDIS measure process and has been in use since 2007, but is currently undergoing re-evaluation.
**The developer tested for construct validity both between the measure components (i.e. the four rates) and a related measure (high-risk medication in those 65 and older). The results suggest that the components are correlated with each other as well as the independent measure of medication safety. In addition, the developer did perform additional analysis to demonstrate that the measure can differentiate between plans.
** Administrative data at the health plan level were utilized in the validity testing. Face validity and construct validity was performed by correlation of results to other measures of medication safety. Face validity utilized a robust process including public comment. Content validity was moderately associated with the High-risk medication in those 65 and older.
** The validity testing level was both measure score and data element. There was both an assessment of face validity and also of construct validity by correlations of the measure with other measures of medication safety. Sample size was adequate in scope.

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: ** Exclusions seem to be determined by the AGS Beer Criteria, not the measure developer.

** The exclusions seem reasonable and appropriate to ensure necessary medications are not restricted or discouraged from patient access. This measure alone provides one piece of evidence on the quality of the provider or health plan. Given the overall high mean incidence and wide variation between health plans, there appears to be considerable differences in practice and quality.

** Exclusions are supported by the evidence.

The performance differences between health plans at the 10th and 90th percentiles are significant.

Criterion 3. Feasibility

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.

- This measure is 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)
- ALL data elements are in defined fields in a combination of electronic sources.
- The measure developer recognizes that despite the clear specifications defined for HEDIS measures, data collection and calculation methods may vary, and other errors may taint the results, diminishing the usefulness of HEDIS data for managed care organization (MCO) comparison.
- In order for the measure to reach its full potential, the developer conducts an independent audit of all HEDIS collection and reporting processes, as well as an audit of the data which are manipulated by those processes, in an effort to verify that HEDIS specifications are met.
- The developer developed a precise, standardized methodology for verifying the integrity of HEDIS collection and calculation processes through a two-part program consisting of an overall information systems capabilities assessment followed by an evaluation of the MCO´s ability to comply with HEDIS specifications. In which certified auditors using standard audit methodologies will help enable purchasers to make more reliable "apples-to-apples" comparisons between health plans.
- In addition to the HEDIS Audit, the developer provides a system to allow “real-time” feedback from measure users. The Policy Clarification Support System receives thousands of inquiries each year on over 100 measures.
- Input from the developer’s auditing and the Policy Clarification Support System informs the annual updating of all HEDIS measures including updating value sets and clarifying the specifications. Measures are re-evaluated on a periodic basis and when there is a significant change in evidence. During re-evaluation information from the auditing and Policy Clarification Support System is used to inform evaluation of the scientific soundness and feasibility of the measure.
- Broad public use and dissemination of these measures is encouraged and the developer has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Use by health care physicians in connection with their own practices is not commercial use. Commercial use of a measure requires the prior written consent of the developer. As used herein, “commercial use” refers to any sale, license or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.

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

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: ** No concerns about feasibility.

** The data elements are routine at the health plan level and adoption of this measure in HEDIS is strong evidence of the feasibility of the measure into practice.

** The measure is generated by and used by healthcare personnel during rendering care. All data elements are available in a combination of electronic sources. The developer verified the integrity of HEDIS collection and calculation processes ad provides a system for real-time feedback for measure users

Criterion 4: Usability and Use

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

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

Current uses of the measure:

Publicly reported? ☒ **Yes** ☐ **No**

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

OR

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

Accountability program details:

- HEALTH PLAN RATINGS/REPORT CARDS: This measure is used to calculate health plan ratings which are reported in Consumer Reports and on the NCQA website. These ratings are based on performance on HEDIS measures among other factors. In 2012, a total of 455 Medicare Advantage health plans, 404 commercial health plans and 136 Medicaid health plans across 50 states were included in the ratings. In 2015 NCQA announced a change in methodology and changed Health Plan Rankings to Health Plan Ratings.
- STATE OF HEALTH CARE ANNUAL REPORT: This measure is publically reported nationally and by geographic regions in the NCQA State of Health Care annual report. This annual report published by NCQA summarizes findings on quality of care. In 2012 the report included measures on 11.5 million Medicare Advantage beneficiaries in 455 Medicare Advantage health plans, 99.4 million members in 404 commercial health plans, and 14.3 million Medicaid beneficiaries in 136 plans across 50 states.
- HEALTH PLAN ACCREDITATION: This measure is used in scoring for accreditation of Medicare Advantage Health Plans. In 2012, a total of 170 Medicare Advantage health plans were accredited using this measure among others covering 7.1 million Medicare beneficiaries. [REPLACE or ADD as appropriate, 336 commercial health plans covering 87 million lives; 77 Medicaid health plans covering 9.1 million lives.] Health plans are scored based on performance compared to benchmarks.
- HEDIS ACCOUNTABLE CARE ORGANIZATION ACCREDITATION: This measure is used in NCQA's ACO Accreditation program, that helps health care organizations demonstrate their ability to improve quality, reduce costs and coordinate patient care. ACO standards and guidelines incorporate whole-person care coordination throughout the health care system.
- HEDIS ACCOUNTABLE CARE ORGANIZATION ACCREDITATION: This measure is used in NCQA's ACO

Accreditation program, that helps health care organizations demonstrate their ability to improve quality, reduce costs and coordinate patient care. ACO standards and guidelines incorporate whole-person care coordination throughout the health care system.

Improvement results:

- Health plan performance rates for the measure have shown slight decreases (i.e., improvement) over the last two years.
- For 2013, 48.5 percent of individuals with a history of falls received at least one high-risk medication and in 2014 this dropped to 48.0 percent. Among individuals with dementia, 49.7 percent received at least one high-risk medication in 2013 and this dropped to 48.5 percent in 2014.
- Among those with chronic kidney disease, 10.4 percent received at least one high-risk medication in 2013 and this dropped to 9.6 percent in 2014.
- Overall, rates from 2013 to 2014 showed a slight decrease, yet the 2014 rates still suggest significant room for improvement, particularly for the history of falls and dementia rates.
- For all rates there is a sizeable gap between the plans at the 10th percentile and 90th percentile, demonstrating a gap in care between the best and worst performing health plans. See section 1b.2 for a summary of recent performance data from health plans.
- Due to recent updates to the medications included in this measure, future rates may show greater room for improvement and variation in performance.

Unexpected findings (positive or negative) during implementation:

- There were no identified unintended consequences for this measure during testing or since implementation.

Potential harms:

- If this measure were to be implemented poorly, there is concern that it could lead to reduced access to medications. There will always be individual cases that will warrant the use of a potentially harmful medication. For example, antidepressants are listed as potentially harmful to patients at risk for falls, however, clinicians should weigh the relative risk of increased falls against the potential benefit of the use of antidepressants for those with severe depression.

Feedback :

- Developer did not identify any specific feedback loops related to this 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

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: ** The measure is in use in several programs such as the Consumer Reports Health Plan ratings. The measure is also used by NCQA to accredit ACOs and physicians.

**It's not clear if the ACO and physician level rates are being publicly reported and it seems that reporting the measure out at this level would be more meaningful to patients.

** Health Plan ratings/report cards; Consumer Reports; NCQA State of Health Care Annual Report; NCQA Health Plan Accreditation; NCQA ACO accreditation
** The measure is used in public reporting and is part of accountability programs -- health plan report cards by NCQA, NCQA annual report, accreditation.
**The only unintended consequence mentioned is an affect on access to needed medication if the measure is not well-implemented. However, each patient should be individually assessed to determine the risk vs. benefit of using a potentially harmful medication.

Criterion 5: Related and Competing Measures

Related or competing measures

- 0022 : Use of High-Risk Medications in the Elderly (DAE)

Harmonization

- This measure is not completely harmonized with 0022. They both have a similar focus (measuring potentially inappropriate medication use in the elderly) and reporting level (health plan), however they have different target populations. This measure targets patients with a specific condition or disease that can experience adverse effects when combined with certain medications that are recommended to be avoided for that condition. NQF 0022 targets a larger population of all older adults and assesses use of high-risk medications that have been recommended to be avoided in all older adults.

Pre-meeting public and member comments

Submitted By: ADVault, Inc.

ADVault believes that people live better lives and, if in a health crisis, can receive better care when they have confidence they can be involved in the creation and implementation of their medical treatment plans and decisions, factors extremely important when it comes to potentially harmful medication being prescribed to the elderly. To do so, they must be able to communicate and express their goals, preferences and priorities for care in a meaningful and actionable way so providers can consider those thoughts. At some point in life, everyone will lose his or her ability to communicate effectively and understand what is being asked of him or her. Healthcare agents should have the confidence to know those value statements as well, in order to fulfill their role as surrogate decision-makers. Non-surrogate family members are comforted with third-party decision-making if they have proof the patient's voice is being heard, clearly understood, and to the extent possible, honored.

Therefore, ADVault strongly recommends providers (1) search for a person's digital emergency, critical and advance care plan (ECACP) upon admission and each time the patient is transitioned to a new site of care, (2) review and update the ECACP in various stages of a person's admission (outpatient or inpatient) and/or illness to ensure respect for the person's goals, preferences and priorities for care, (3) link the digital ECACP to the EHR and/or patient portal in order to ease access and address security, privacy and patient consent concerns, (4) track and make available the number of ECACPs found, opened and re-visited, and the impact they have on the care of the patient, as well as patient, family and caregiver satisfaction, such data to be reported in a manner such that: (a) consumers can make better choices about hospitals and doctors; (b) doctors improve the satisfaction and quality of their work; and (c) hospital administrators gauge performance and align caregiving goals with actual outcomes. Finally, if no ECACP can be found via standards-based healthcare IT transport mechanisms, the hospital/provider should engage the patient to create one whenever possible.

Submitted by: Centers for Disease Control and Prevention

CDC strongly supports a patient safety measure related to medication management in older adults; however, we are concerned that the CDC data cited is not appropriately applied and the measure may not efficiently reduce adverse drug events (ADEs). First, the measure rationale is that reduction in "high-risk medication" (HRM) use

"should decrease morbidity and mortality" associated with ADEs and CDC data are cited in the discussion of measure impact. However, CDC data indicate the opposite--Beers Criteria (BC) HRMs are not leading causes of emergency department (ED) visits or hospitalizations for ADEs (Ann Intern Med 2007;147:755-65; N Engl J Med 2011;365:2002-12). Approximately 1% of U.S. hospitalizations for ADEs among older adults involve BC HRMs, while approximately 66% involve 3 other drug classes (warfarin, antidiabetics, oral antiplatelets). After accounting for prescribing, the hospitalizations rate for ADEs from these 3 drug classes is at least 40 times higher than the hospitalization rate for ADEs from BC HRMs (N Engl J Med 2011;365:2002-12).

Second, although there are a few studies to support an epidemiologic association of BC HRMs with health outcomes, there are many other studies that do not support this finding. The studies cited in the measure are based on older BC versions. We are not aware of new data demonstrating that use of the updated BC is associated with morbidity, mortality, or resource utilization reductions. Third, using a composite measure targeting hundreds of drugs/interactions obscures the contribution of specific drugs and thus cannot be efficiently used to implement interventions (J Hosp Med 2008;3:87-90). One-half of Medicare Advantage beneficiaries meet criteria for HRM drug-disease interactions, suggesting the measure is not useful for targeting the highest risk drugs. Fourth, basing a broad healthcare quality measure on the "potentially inappropriate" concept is problematic because it supersedes the treating clinician's judgment without having supporting information for that clinical judgment. The 2015 BC update states: "these criteria are not meant to be applied in a punitive manner. Prescribing decisions are not always clear-cut, and clinicians must consider multiple factors...Quality measures must be...measured with limited information and thus...cannot perfectly distinguish appropriate from inappropriate care". The BC is a useful tool to guide individual clinical decisions; however, as a quality measure, it is likely to have minimal population impact. A fundamental criterion of NQF measures is that they be aligned with national health priorities; for medication safety, these have been defined as improving safe use of anticoagulants, antidiabetics, and opioids (health.gov/hcq/ade-action-plan.asp). Incorporation of these medications into national quality measures will go further toward improving health outcomes for older Americans than measures focused on HRMs.

Submission materials attachments…

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

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

Measure Title: Potentially Harmful Drug-Disease Interactions in the Elderly

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

Date of Submission: Click here to enter a date

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with

patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

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

Outcome

☐ Health outcome: Click here to name the health outcome

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

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

☒ Process: Prescribing of potentially harmful drugs for the elderly

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

Clinician assesses the patient's underlying diseases/conditions that put them at higher risk for adverse drug events. Clinician weighs risks and benefits of prescribing medications recommended to be avoided for the patient's disease/condition

↓

Measured Process: Clinician judiciously prescribes potentially harmful medications, selecting alternative pharmacologic and non-pharmacologic treatment approaches when possible

↓

Adverse drug events are avoided for the

↓

Desired Outcome:
Morbidity and mortality is

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

American Geriatrics Society 2015 Beers Criteria Update Expert Panel. 2015. American Geriatrics Society 2015 Updated Beers Criteria for Potentially Inappropriate Medication Use in Older Adults. Journal of the American Geriatrics Society, 63(11): 2227-2246. Guideline available at: http://geriatricscareonline.org/ProductAbstract/american-geriatrics-society-updated-beers-criteria-for-potentially-inappropriate-medication-use-in-older-adults/CL001

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

Language in the below table is taken verbatim from Table 3 (pages 15-18) of the *American Geriatrics Society (AGS) 2015 Updated Beers Criteria for Potentially Inappropriate Medication Use in Older Adults*.

History of Falls or Fractures (page16-17)				
Drugs	Rationale	Recommendation	Quality of Evidence	Strength of Recommendation
Anticonvulsants Antipsychotics	May cause ataxia, impaired psychomotor function, syncope,	Avoid unless safer alternatives are	High	Strong

Benzodiazepines Nonbenzodiazepine, benzodiazepine receptor agonist hypnotics -Eszopiclone -Zaleplon -Zolpidem TCAs SSRIs Opioids[1]	additional falls; shorter-acting benzodiazepines are not safer than long-acting ones If one of the drugs must be used, consider reducing use of other CNS-active medications that increase risk of falls and fractures (i.e., anticonvulsants, opioidreceptor agonists, antipsychotics, antidepressants, benzodiazepine receptor agonists, other sedatives and hypnotics) and implement other strategies to reduce fall risk	not available; Avoid anticonvulsants except for seizure and mood disorders[1] Opioids: avoid, excludes pain management due to recent fractures or joint Replacement[2]	Opioids: moderate	Opioids: strong

[1]Anticonvulsants are included in the measure because the conditions for which there is appropriate use can be reliably identified using claims data, so we can exclude those with appropriate use (see Guiding Principles under section 1a.7.1).

[2]Opioids are not included in the measure due to the caveat in the recommendation statement that opioid use for pain management due to recent fractures or joint replacement is appropriate. These uses cannot be reliably identified using claims data alone, so we cannot exclude those with appropriate use (see Guiding Principles under section 1a.7.1).

Chronic kidney disease Stages IV or less (creatinine clearance <30 mL/min) (page 18)				
Drugs	Rationale	Recommendation	Quality of Evidence	Strength of Recommendation
NSAIDs (non-COX and COX-selective, oral and parenteral)	May increase risk of acute kidney injury and further decline of renal function	Avoid	Moderate	Strong

Dementia or cognitive impairment (page 15-16)				
Drugs	Rationale	Recommendation	Quality of Evidence	Strength of Recommendation
Anticholinergics (see Table 7 for full list) Benzodiazepines H2-receptor antagonists Nonbenzodiazepine,	Avoid because of adverse CNS Effects Avoid antipsychotics for behavioral problems of dementia or delirium unless	Avoid	Moderate	Strong

benzodiazepine receptor agonist hypnotics -Eszopiclone -Zolpidem -Zaleplon Antipsychotics, chronic and as-needed use	nonpharmacological options (e.g., behavioral interventions) have failed or are not possible and the older adult is threatening substantial harm to self or others. Antipsychotics are associated with greater risk of cerebrovascular accident (stroke) and mortality in persons with dementia			

Table 7: Drugs with Strong Anticholinergic Properties

Antihistamines Brompheniramine Carbinoxamine Chlorpheniramine Clemastine Cyproheptadine Dexbrompheniramine Dexchlorpheniramine Dimenhydrinate Diphenhydramine (oral) Doxylamine Hydroxyzine Meclizine Triprolidine	Antiparkinsonian agents Benztropine Trihexyphenidyl	Skeletal muscle relaxants Cyclobenzaprine Orphenadrine	Antidepressants Amitriptyline Amoxapine Clomipramine Desipramine Doxepin (>6 mg) Imipramine Nortriptyline Paroxetine Protriptyline Trimipramine	Antiemetic Prochlorperazine Promethazine
Antipsychotics Chlorpromazine Clozapine Loxapine Olanzapine Perphenazine Thioridazine Trifluoperazine	Antiarrhythmic Disopyramide	Antimuscarinics (urinary incontinence) Darifenacin Fesoterodine Flavoxate Oxybutynin Solifenacin Tolterodine Trospium	Antispasmodics Atropine (excludes ophthalmic) Belladonna alkaloids Clidiniumchlordiazepoxide Dicyclomine Homatropine (excludes ophthalmic) Hyoscyamine	

			Propantheline Scopolamine (excludes ophthalmic)	

1a.4.3. Grade assigned to the quoted recommendation <u>with definition</u> of the grade:

Quality of Evidence	
High	Evidence includes consistent results from well designed, well-conducted studies in representative populations that directly assess effects on health outcomes (≥2 consistent, higher-quality randomized controlled trials or multiple, consistent observational studies with no significant methodological flaws showing large effects)
Moderate	Evidence is sufficient to determine risks of adverse outcomes, but the number, quality, size, or consistency of included studies; generalizability to routine practice; or indirect nature of the evidence on health outcomes (≥1 higher-quality trial with >100 participants; ≥2 higher-quality trials with some inconsistency; ≥2 consistent, lower-quality trials; or multiple, consistent observational studies with no significant methodological flaws showing at least moderate effects) limits the strength of the evidence
Strength of Recommendation	
Strong	Benefits clearly outweigh harms, adverse events, and risks, or harms, adverse events, and risks clearly outweigh benefits

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

Quality of Evidence	
Low	Evidence is insufficient to assess harms or risks in health outcomes because of limited number or power of studies, large and unexplained inconsistency between higher-quality studies, important flaws in study design or conduct, gaps in the chain of evidence, or lack of information on important health outcomes
Strength of Recommendation	
Weak	Benefits may not outweigh harms, adverse events, and risks
Insufficient	Evidence inadequate to determine net harms, adverse events, and risks

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

Qaseem A, Snow V, Owens DK et al. The development of clinical practice guidelines and guidance statements of the American College of Physicians: Summary of methods. Ann Intern Med 2010;153:194–199.

The GRADE working group. GRADE guidelines—best practices using the GRADE framework. Journal of Clinical Epidemiology [on-line]. Available at http://www.gradeworkinggroup.org/publications/jce_series.htm

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☒ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

This measure assesses if patients with specific conditions/diseases received medications that are potentially harmful for their condition. This measure is based on the AGS Beers Criteria recommendations against the use of potentially harmful medications in older adults with specific conditions (i.e., recommendations contained in Table 3), where the potential harms of the medication outweigh the benefits. The AGS Beers Criteria identifies 12 conditions where there are potentially inappropriate medications. This measure includes three of those conditions: history of falls or fracture, dementia or cognitive impairment, and chronic kidney disease. Below are the guiding principles that were developed to determine which conditions from the evidence would be included in the measure and which medications would be included in the measure.

Guiding Principles

Include conditions and medications listed in *Table 3: 2015 AGS Beers Criteria for Potentially Inappropriate Medication Use in Older Adults Due to Drug-Disease or Drug-Syndrome Interactions That May Exacerbate the Disease or Syndrome*.

The following criteria were used to determine which conditions from Table 3 should be included in the measure:

1. Do not include conditions where all potentially harmful medications are already listed in Table 2 of the Beers Criteria.
2. Do not include conditions that are rare and would not provide a sufficient denominator count for quality measurement.
3. Only include conditions where the performance rate indicates there is room for improvement (i.e., greater than minimum use of the potentially inappropriate medication).
4. Only include conditions that can be reliably identified by claims data.
5. Do not include conditions where all potentially harmful medications are primarily available over the counter.

The following criteria were used to determine which medications from Table 3 should be included in the measure:

1. Include only prescription medications.
2. Include only medications with strong recommendations to avoid.
3. When a caveat is listed in Table 3 as an appropriate use of the medication and can be identified in claims, add the medication with an exclusion for the identifiable caveat. For example, anticonvulsants should be avoided except for those with seizure disorders; therefore, there is an exclusion for seizure disorders in the History of Falls rate.
4. When a caveat is listed for a medication class that cannot be identified in claims data, the medication (class) may be included in the measure if the non-identifiable caveat is considered rare. For example, the caveat for antipsychotics for people with dementia is that they should be avoided unless nonpharmacological options have failed and the patient is a threat to self or others. This caveat would be a rare event that cannot be identified in claims.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

The grade assigned by AGS to the quality of evidence and the strength of the recommendations varied by each recommendation. See table under 1a.4.2 for the quality of evidence and strength of recommendation grades given to each recommendation.

AGS Quality of Evidence Definitions:

High: Evidence includes consistent results from well designed, well-conducted studies in representative populations that directly assess effects on health outcomes (≥2 consistent, higher-quality randomized controlled trials or multiple, consistent observational studies with no significant methodological flaws showing large effects)

Moderate: Evidence is sufficient to determine risks of adverse outcomes, but the number, quality, size, or consistency of included studies; generalizability to routine practice; or indirect nature of the evidence on health outcomes (≥1 higher-quality trial with >100 participants; ≥2 higher-quality trials with some inconsistency; ≥2 consistent, lower-quality trials; or multiple,

consistent observational studies with no significant methodological flaws showing at least moderate effects) limits the strength of the evidence

AGS Strength of Recommendation Definitions:

Strong: Benefits clearly outweigh harms, adverse events, and risks, or harms, adverse events, and risks clearly outweigh benefits

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

AGS Quality of Evidence Definitions:

Low: Evidence is insufficient to assess harms or risks in health outcomes because of limited number or power of studies, large and unexplained inconsistency between higher-quality studies, important flaws in study design or conduct, gaps in the chain of evidence, or lack of information on important health outcomes

AGS Strength of Recommendation Definitions:

Weak: Benefits may not outweigh harms, adverse events, and risks

Insufficient: Evidence inadequate to determine net harms, adverse events, and risks

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: 2004-2014

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

The Beers Criteria were first published in 1991. Since that time the criteria have been regularly updated based off of the existing criteria and any new evidence published since the last update. The American Geriatrics Society forms an expert panel to update the Beers Criteria every few years. The panel works from the previous evidence review and then reviews any new evidence published since that last review to update the recommendations in the Beers Criteria. The 2015 review by the AGS 2015 Beers Criteria Update Expert Panel included review of 60 systematic reviews and meta analyses, 49 randomized control trials (RTCs) and 233 observational studies and other types of publications.

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

Overall the quality of the evidence for each of the conditions/diseases included in quality measure is good. The dementia and chronic kidney disease section have moderate evidence and the history of falls section has high evidence. In addition to conducting a systematic review of the evidence, the AGS 2015 Beers Criteria Update Expert Panel also used technical experts and a public comment period for additional validity.

History of falls: Evidence for the recommendation to avoid certain medications (anticonvulsants, antipsychotics, benzodiazepine and nonbenzodiazepine hypnotics, tricyclic antidepressants and SSRIs) for individuals with a history of falls was rated as high quality. It includes 5 systematic reviews of evidence in addition to cohort studies.

Anticonvulsants: 2 systematic review, 1 cohort

Antipsychotics: 2 systematic review; 3 cohort; 1 case-control

Benzodiazepines: 3 systematic review; 2 cohort; 3 case-control

Nonbenzodiazepine hypnotics: 2 systematic review; 1 cohort; 1 case-control

Tricyclic Antidepressants: 3 systematic review; 3 cohort; 1 case-control

SSRIs: 2 systematic review; 3 cohort; 1 case-control

Dementia or cognitive impairment: Evidence for the recommendation to avoid certain medications (anticholinergic drugs, benzodiazepines, H2-receptor antagonists, benzodiazepine and nonbenzodiazepine hypnotics, antipsychotics) for individuals with dementia was rated as moderate quality. It includes 2 systematic reviews and 3 randomized control studies in addition to cohort studies.

Antiemetics:1 cohort

Antipsychotics: 6 cohort

Benzodiazepines: 5 cohort

Nonbenzodiazepine hypnotics: 2 cohort

Tricyclic Antidepressants: 1 systematic review; 3 cohort

H2 Receptor Antagonists: 2 cohort

Antihistamines: 1 cohort

Antispasmodics: 1 cohort

Antimuscaninics (oral): 1 cohort

Parkinson agents: 1 cohort

Skeletal muscle relaxants: 1 cohort

SSRIs: 1 systematic review; 2 cohort

Antiarrhythmic: 1 cohort

Chronic kidney disease: Evidence for the recommendation to avoid NSAIDs for individuals with chronic kidney disease (stages IV or less [creatinine clearance <30 mL/min)]) was rated as moderate quality. It includes 1 randomized control study and 5 cohort studies.

Cox-2 Selective NSAIDs: 1 randomized control study; 1 cohort

Nonaspirin NSAIDs: 1 randomized control study; 4 cohort

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

Recommendations in the Beers criteria are based on studies that explain the rationale for why a medication group is potentially harmful for a patient with a certain condition. Below is a summary of the number and types of studies supporting the recommendation for each drug class. Summaries of each study can be found on the American Geriatrics Society's website: http://www.americangeriatrics.org/.

Condition	Class of Drugs	Studies that support recommendation	Recommendation
Dementia or cognitive impairment	Anticholinergics (see Table 7 for full list) Benzodiazepines H2-receptor antagonists Nonbenzodiazepine, benzodiazepine receptor agonist hypnotics Eszopiclone Zolpidem	2015 Criteria: Chavant 2011 Kalicsh Ellet 2014 From previous criteria: Boustani 2007 Hanlon2004 Finkle 2011 Frey 2011 Paterniti 2002 Rasmussen 1999 Rudolph 2008 Schneider 2005 Schneider 2006a Schneider 2006b Seitz 2011 Vigen 2011 Wright 2009	Avoid because of adverse CNS Effects Avoid antipsychotics for behavioral problems of dementia or delirium unless nonpharmacological options (e.g., behavioral interventions) have failed or are not possible and the older adult is threatening substantial harm to self or others. Antipsychotics are associated with greater risk of cerebrovascular accident (stroke) and mortality in persons with dementia
History of falls or fractures	Anticonvulsants Antipsychotics Benzodiazepines Nonbenzodiazepine, benzodiazepine receptor agonist hypnotics -Eszopiclone -Zaleplon -Zolpidem	Rolita 2013 Soderberg 2013 From previous criteria: Allain 2005 Berdot 2009 Deandrea 2010 Ensrud 2003 Hartikainen	May cause ataxia, impaired psychomotor function, syncope, additional falls; shorter-acting benzodiazepines are not safer than long-acting ones If one of the drugs must be used, consider reducing use of other CNS-active medications that increase risk of falls and fractures (i.e., anticonvulsants, opioidreceptor agonists, antipsychotics, antidepressants, benzodiazepine receptor agonists, other sedatives and hypnotics) and implement other strategies to

	TCAs SSRIs	2007 Jalbert 2010 Liperoti 2007 Mets 2010 Sterke 2008 Turner 2011 van der Hooft 2008 Vestergaard 2008 Wagner 2004 Wang 2001a Wang 2001b	reduce fall risk
Chronic Kidney disease	NSAIDs (non-COX and COX-selective, oral and parenteral)	Gooch 2007 Griffin 2000 Lafrance 2009 Murray 1995 Schneider 2006 Winkelmayer 2008	May increase risk of acute kidney injury and further decline of renal function

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

As part of their review of the evidence, the AGS 2015 Beers Criteria Update Expert Panel identified subgroups of patients who should be exempt from the criteria and for whom listed medications may be appropriate. In addition, a patient could have a condition or comorbidity that would merit the use of a medication on the list, even if the comorbidity is not specifically listed in the criteria. The panel noted that exclusions to the criteria should not be expanded to include all adults 65 and older when only a portion of individuals may benefit from use of these medications. The criteria are designed to assist providers in the prescribing of potentially harmful medications, and should not be taken as strict criteria to avoid use in all patients without weighing the harms and benefits for individual cases.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

To our knowledge there have been no published studies since the systematic review that would impact the recommendations.

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

Measure Information

This document contains the information submitted by measure developers/stewards, but is organized according to NQF's measure evaluation criteria and process. The item numbers refer to those in the submission form but may be in a slightly different order here. In general, the item numbers also reference the related criteria (e.g., item 1b.1 relates to subcriterion 1b).

Brief Measure Information

NQF #: 2993
De.2. **Measure Title:** Potentially Harmful Drug-Disease Interactions in the Elderly
Co.1.1. **Measure Steward:** National Committee for Quality Assurance
De.3. **Brief Description of Measure:** The percentage of patients 65 years of age and older who have evidence of an underlying disease, condition or health concern and who are dispensed an ambulatory prescription for a potentially harmful medication, concurrent with or after the diagnosis. Four rates are reported for this measure:

-Rate 1: The percentage of those with a history of falls that received a potentially harmful medication
-Rate 2: The percentage of those with dementia that received a potentially harmful medication
-Rate 3: The percentage of those with chronic kidney disease that received a potentially harmful medication
-Rate 4: Total rate

A lower rate represents better performance for all rates.
1b.1. **Developer Rationale:** Lowering the rate of potentially harmful drug-disease interactions in the elderly population should decrease morbidity and mortality associated with adverse drug reactions.

S.4. **Numerator Statement:** Numerator 1: Patients with a history of falls who received at least one potentially harmful medication from Table DDE-A or Table DDE-B
Numerator 2: Patients with a diagnosis of dementia who received at least one potentially harmful medication from Table DDE-D
Numerator 3: Patients with chronic kidney disease who received at least one potentially harmful medication from Table DDE-E
Numerator 4: The sum of the three numerators
S.7. **Denominator Statement:** All patients ages 65 years of age and older with a history of falls, dementia or chronic kidney disease in the measurement year or the year prior to the measurement year.
S.10. **Denominator Exclusions:** The following are exclusions for the condition-specific rates and total rate:

For those who meet denominator criteria for the history of falls rate (Rate 1): exclude those with a diagnosis of psychosis, schizophrenia, bipolar disorder or seizure disorder.

For those who meet denominator criteria for those with dementia rate (Rate 2): exclude those with a diagnosis of psychosis, schizophrenia or bipolar disorder.

De.1. **Measure Type:** Process
S.23. **Data Source:** Administrative claims, Electronic Clinical Data, Electronic Clinical Data : Pharmacy
S.26. **Level of Analysis:** Health Plan, Integrated Delivery System

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, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

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

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., the benefits or improvements in quality envisioned by use of this measure)*
Lowering the rate of potentially harmful drug-disease interactions in the elderly population should decrease morbidity and mortality associated with adverse drug reactions.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
The following data are extracted from HEDIS data collection for Medicare Advantage Health Plans (including both HMO and PPO plans). Performance data is summarized at the health plan level and summarized by mean, standard deviation, and performance at the 10th, 25th, 50th, 75th and 90th percentile. Data is stratified by year.

The following data demonstrate variation in all four rates of the measure. For 2014, 48.0 percent of individuals with a history of falls received at least one high-risk medication. Among individuals with dementia, 48.5 percent received at least one high-risk medication and among those with chronic kidney disease, 9.6 percent received at least one high-risk medication. The national mean performance for the total rate was 41.5 percent. Overall, rates from 2013 to 2014 showed a slight decrease, yet the 2014 rates still suggest significant room for improvement, particularly for the history of falls and dementia rates. For all rates there is a sizeable gap between the plans at the 10th percentile and 90th percentile, demonstrating a gap in care between the best and worst performing health plans.

Rate 1 (History of Falls)
YEAR| N | MEAN | ST DEV | 10TH (Better) | 25TH | 50TH | 75TH | 90TH (Worse)| Interquartile Range
2013 | 412 | 48.5% | 8.4 | 38.6% | 43.3% | 47.9% | 53.7% | 58.4 | 10.4
2014*| 387 | 48.0% | 8.3 | 38.8% | 43.1% | 47.7% | 51.9% | 58.5 | 8.8
*For 2014 the average eligible population was 1,411, with a standard deviation of 2,775

Rate 2 (Dementia)
YEAR| N| MEAN | ST DEV | 10TH (Better) | 25TH | 50TH | 75TH | 90TH (Worse) | Interquartile Range
2013 | 417 | 49.7% | 9.2 | 40.1% | 43.4% | 48.5% | 54.7% | 61.6% | 11.3
2014*| 385 | 48.5% | 9.1 | 39.2% | 42.8% | 46.8% | 52.8% | 61.0% | 10.0
*For 2014 the average eligible population was 1,330, with a standard deviation of 2,395

Rate 3 (Chronic Kidney Disease)
YEAR| N| MEAN | ST DEV | 10TH (Better) | 25TH | 50TH | 75TH | 90TH (Worse) | Interquartile Range
2013 | 379 | 10.4% | 6.2 | 4.6% | 6.4% | 8.9% | 12.9% | 18.6% | 6.5
2014*| 356 | 9.6% | 6.1 | 3.9% | 5.8% | 8.1% | 12.0% | 17.1% | 6.2
*For 2014 the average eligible population was 661, with a standard deviation of 1,200

Rate 4 (Total)
YEAR| N| MEAN | ST DEV | 10TH (Better) | 25TH | 50TH | 75TH | 90TH (Worse) | Interquartile Range
2013 | 437 | 42.4% | 8.0 | 33.5% | 37.1% | 40.9% | 46.5% | 52.8% | 9.4
2014*| 412 | 41.5% | 7.9 | 33.5% | 36.7% | 40.3% | 44.9% | 51.3% | 8.2
*For 2014 the average eligible population was 3,144, with a standard deviation of 6,055

The data referenced are extracted from HEDIS data collection reflecting the most recent years of measurement for this measure. In 2014, HEDIS measures covered more than 171 million people from 814 HMO health plans and 353 PPO health plans. Below is a description of the denominator for this measure. It includes the number of health plans reporting the measure and the mean eligible population for the measure across health plans.

Rate 1 (History of Falls)
YEAR | N Plans | Mean Denominator Size per plan
2013 | 412 | 1,165
2014 | 387 | 1,411

Rate 2 (Dementia)
YEAR | N Plans | Mean Denominator Size per plan
2013 | 417 | 1,156
2014 | 385 | 1,330

Rate 3 (Chronic Kidney Disease)
YEAR | N Plans | Mean Denominator Size per plan
2012 | 305 | 268
2013 | 379 | 558
2014 | 356 | 661

Rate 4 (Total)
YEAR | N Plans | Mean Denominator Size per plan
2013 | 437 | 2,689
2014 | 412 | 3,144

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.
Section 1b.2 references data from the most recent two years of measurement for this measure. The data in section 1b.2 includes percentiles, mean, interquartile range and standard deviation and demonstrates room for improvement.

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*
NCQA does not currently collect performance data stratified by race, ethnicity, or language. Escarce et al. have described in detail the difficulty of collecting valid data on race, ethnicity and language at the health plan level (Escarce, 2011). While not specified in the measure, this measure can also be stratified by demographic variables, such as race/ethnicity or socioeconomic status, in order to assess the presence of health care disparities. The HEDIS Health Plan Measure Set contains two measures that can assist with stratification to assess health care disparities. The Race/Ethnicity Diversity of Membership and the Language Diversity of Membership measures were designed to promote standardized methods for collecting these data and follow Office of Management and Budget and Institute of Medicine guidelines for collecting and categorizing race/ethnicity and language data. In addition, NCQA's Multicultural Health Care Distinction Program outlines standards for collecting, storing and using race/ethnicity and language data to

assess health care disparities. Based on extensive work by NCQA to understand how to promote culturally and linguistically appropriate services among plans and providers, we have many examples of how health plans have used HEDIS measures to design quality improvement programs to decrease disparities in care.
Escare J.J., Carreon R., Vesolovskiy G., and Lawson E.H. 2011. Collection Of Race And Ethnicity Data By Health Plans Has Grown Substantially, But Opportunities Remain To Expand Efforts. Health Affairs 20(10): 1984-1991.
Centers for Disease Control and Prevention. 2010. Vital Signs. http://www.cdc.gov/VitalSigns/pdf/2010-07-vitalsigns.pdf (Accessed July 8, 2011).

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
While disparities for this measure have not been well studied, there is some evidence to suggest that women are more likely to receive a potentially inappropriate medication than men. A retrospective cohort study of 966,000 men and women treated by the Veteran's Health Administration showed that women were more likely than men to receive medications that may have harmful interactions with chronic conditions as described by the Beers Criteria (Bierman et al., 2007). In a different study, a retrospective database analysis of HEDIS data from the Department of Veterans Affairs found that Hispanics and those with no copayments had higher rates of medications listed as potentially harmful than whites or those with required copayments (Pugh, 2011).
Bierman, A.S., M.J.V. Pugh, I. Dhalla, M. Amuan, B.G. Fincke, A. Rosen, D.R. Berlowitz. 2007. "Sex differences in inappropriate prescribing among elderly veterans." The American Journal of Geriatric Pharmacotherapy, 5(2):147-161.
Pugh, Mary Jo V., et al. "Exposure to Potentially Harmful Drug–Disease Interactions in Older Community-Dwelling Veterans Based on the Healthcare Effectiveness Data and Information Set Quality Measure: Who Is at Risk?." Journal of the American Geriatrics Society 59.9 (2011): 1673-1678.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, A leading cause of morbidity/mortality, High resource use, Patient/societal consequences of poor quality, Severity of illness
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
There is clinical consensus that in the elderly certain medications are associated with increased risk of harm from drug side-effects and drug toxicity; these medications pose a concern for patient safety. Use of potentially inappropriate medications (PIMs) in the elderly can lead to poor health outcomes including adverse drug events, confusion, falls, hospitalizations and even death. Despite widely-accepted medical consensus that certain drugs increase the risk of harm to the elderly and should generally be avoided, these drugs are still frequently prescribed to the elderly. In a study of health outcomes, 40% of individuals 65 and older filled at least one PIM and 13% filled two or more (Fick et al. 2008). In this population, 14.3% of those who had at least one PIM had a drug-related problem, whereas only 4.7% of those with no PIMs had a drug-related problem. PIM use in the elderly has been connected to increased hospitalization and increased risk of death (Lau et al., 2004). Preventing poor health effects from use of PIMs is expected to be a growing concern with the increasing population of adults over 65, longer life expectancies and the introduction of new medications (Rothberg et al., 2008).

Reducing use of PIMs in the elderly also represents an opportunity to reduce the costs associated with harm from medications (e.g., hospitalizations from drug toxicity) and encourage clinicians to consider alternative, safer medications. Conservative estimates of extra costs due to potentially inappropriate medications in the elderly average $7.2 billion a year (Fu, 2007). The annual direct costs of preventable ADEs in the Medicare population have been estimated to exceed $800 million (Institute of Medicine, 2007). Reducing unnecessary prescribing will also help to reduce cost, given that the elderly population represent one third of all prescription drug expenditures in the U.S. but comprises only 13 percent of the population (Families USA, 2000). While expenditures for prescription drugs in the US are disproportionately clustered among those 65 years and older, this population is twice as likely as those below age 65 to experience adverse drug events and is almost seven times as likely to be hospitalized for adverse drug events (Budnitz, 2006).

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

Budnitz, D., D.A. Pollock, K.N. Widenbach, A.B. Mendelson, T.J. Schroeder, and J.L. Annest. 2006. "National Surveillance of Emergency Department Visits for Outpatient Adverse Drug Events." Journal of the American Medical Association 296:1858-1866.

Families USA, Cost Overdose: Growth in Drug Spending for the Elderly, 1992-2010. 2000. Washington, DC: Families USA. July, p. 2.

Fick, D.M., L.C. Mion, M.H. Beers, J.L. Waller. 2008. "Health Outcomes Associated with Potentially Inappropriate Medication Use in Older Adults." Research in Nursing & Health. 31(1): 42-51.

Fick, D.M., and T.P. Selma. 2012. 2012 American Geriatrics Society Beers Criteria: New Year, New Criteria, New Perspective. The American Geriatrics Society.

Fu, A.Z., J.Z. Jiang, J.H. Reeves, J.E. Funcham, G.G. Liu, M. Perri. 2007. "Potentially Inappropriate Medication Use and Healthcare Expenditures in the US Community-Dwelling Elderly." Medical Care 45: 472-6.

Institute of Medicine (IOM). 2007. Preventing Medication Errors/Committee on Identifying and Preventing Medication Errors. Ed. Aspden P., J.A. Wolcott, J.L. Bootman, L.R. Cronenwatt LR. Quality Chasm Series. Washington, DC: National Academy Press.
Lau, D.T.. J.D. Kasper, D.E. Potter, A. Lyles. 2004 "Potentially Inappropriate Medication Prescriptions Among Elderly Nursing Home Residents: Their Scope and Associated Resident and Facility Characteristics." Health Services Research 39(5): 1257-1276.
Rothberg, M.B., P.S. Perkow, F. Liu, B. Korc-Grodzicki, M.J. Brennan, S. Bellantonio, M. Heelon, P.K. Lindenauer. 2008. "Potentially Inappropriate Medication Use in Hospitalized Elders." Journal of Hospital Medicine. 3: 91-102.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

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 subcriteria 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):*

Prevention

De.6. **Cross Cutting Areas** *(check all the areas that apply):*

Safety : Medication Safety

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

N/A

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

This is not an eMeasure **Attachment:**

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

Attachment **Attachment:** DDE_Value_Sets-635979522717911582.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
N/A

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Numerator 1: Patients with a history of falls who received at least one potentially harmful medication from Table DDE-A or Table DDE-B
Numerator 2: Patients with a diagnosis of dementia who received at least one potentially harmful medication from Table DDE-D
Numerator 3: Patients with chronic kidney disease who received at least one potentially harmful medication from Table DDE-E
Numerator 4: The sum of the three numerators

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
January 1 of the year prior to the measurement year through December 31 of the measurement year (24-month period).

S.6. **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, 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.
Rate 1 numerator: Dispensed an ambulatory prescription for an anticonvulsant, nonbenzodiazepine hypnotic, or SSRI (Table DDE-A), antipsychotic, benzodiazepine, nonbenzodiazepine hypnotic or tricyclic antidepressant (Table DDE-B) on or between the index episode start data and December 31 of the measurement year.

Rate 2 numerator: Dispensed an ambulatory prescription for an antipsychotic, benzodiazepine, nonbenzodiazepine hypnotic or tricyclic antidepressant (Table DDE-B), or H2 receptor antagonist or anticholinergic agent (Table DDE-D) on or between the IESD and December 31 of the measurement year.

Rate 3 numerator: Dispensed an ambulatory prescription for an NSAID or Cox-2 selective NSAID (Table DDE-E) on or between the IESD and December 31 of the measurement year.

Rate 4 numerator: The sum of numerators 1, 2 and 3.

Note: Do not include denied claims.

…
Table DDE-A: Potentially Harmful Drugs – Rate 1
Anticonvulsants:
Carbamazepine, Clobazam, Divalproex sodium, Ethosuximide, Ethotoin, Ezogabine, Felbamate, Fosphenytoin, Gabapentin, Lacosamide, Lamotrigine, Levetiracetam, Mephobarbital, Methsuximide, Oxcarbazepine, Phenobarbital, Phenytoin, Pregabalin, Primidone, Rufinamide, Tiagabine HCL, Topiramate, Valproate sodium, Valproic acid, Vigabatrin, Zonisamide

SSRIs:
Citalopram, Escitalopram, Fluoxetine, Fluvoxamine, Paroxetine, Setraline

---
Table DDE-B: Potentially Harmful Drugs – Rate 1 (History of Falls) and Rate 2 (Dementia)
Antipsychotics:
Aripiprazole, Asenapine, Brexpiprazole, Cariprazine, Chlorpromazine, Clozapine, Fluphenazine, Haloperidol, Iloperidone, Loxapine, Lurasidone, Molindone, Olanzapine, Paliperidone, Perphenazine, Pimozide, Quetiapine, Risperidone, Thioridazine, Thiothixene, Trifluoperazine, Ziprasidone

Benzodiazepine hypnotics:
Alprazolam, Chlordiazepoxide products, Clonazepam, Clorazepate-Dipotassium, Diazepam, Estazolam, Flurazepam HCL, Lorazepam, Midazolam HCL, Oxazepam, Quazepam, Temazepam, Triazolam

Nonbenzodiazepine hypnotics:
Eszopiclone, Zaleplon, Zolpidem

Tricyclic antidepressants:
Amitriptyline, Amoxapine, Clomipramine, Desipramine, Doxepin (>6 mg), Imipramine, Nortriptyline, Protriptyline, Trimipramine

---
Table DDE-D: Potentially Harmful Drugs – Rate 2 (Dementia)
H2 receptor antagonists:
Cimetidine, Famotidine, Nizatidine, Ranitidine

Anticholinergic agents, antiemetics:
Prochlorperazine, Promethazine

Anticholinergic agents, antihistamines:
Carbinoxamine, Chlorpheniramine, Hydroxyzine products, Brompheniramine, Clemastine, Cyproheptadine, Promethazine, Triprolidine, Dimenhydrinate, Diphenhydramine, Meclizine, Dexbrompheniramine, Dexchlorpheniramine, Doxylamine

Anticholinergic Agents, antimuscarinics (oral)
Atropine, Homatropine, Belladonna alkaloids, Dicyclomine, Hyoscyamine, Propantheline, Scopolamine, Clidinium-chlordiazepoxide

Anticholinergic agents, antimuscarinics (oral)
Darifenacin, Fesoterodine, Solifenacin, Trospium, Flavoxate, Oxybutynin, Tolterodine

Anticholinergic agents, anti-Parkinson agents
Benztropine, Trihexyphernidyl

Anticholinergic agents, skeletal muscle relaxants
Cyclobenzaprine, Orphenadrine

Anticholinergic agents, SSRIs:
Paroxetine

Anticholinergic agents, antiarrhythmic:
Disopyramide

Table DDE-E: Cox-2 Selective NSAIDs and Nonasprin NSAIDs
Cox-2 Selective NSAIDs:
Celecoxib

Nonaspirin NSAIDs:
Diclofenac potassium, Diclofenac sodium, Etodolac, Fenoprofen, Flurbiprofen, Ibuprofen, Indomethacin, Ketoprofen, Ketorolac, Meclofenamate, Mefenamic acid, Meloxicam, Nabumetone, Naproxen, Naproxen sodium, Oxaprozin, Piroxicam, Sulindac, Tolmetin

S.7. Denominator Statement *(Brief, narrative description of the target population being measured)*
All patients ages 65 years of age and older with a history of falls, dementia or chronic kidney disease in the measurement year or the year prior to the measurement year.

S.8. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, Populations at Risk : Dual eligible beneficiaries, Populations at Risk : Individuals with multiple chronic conditions, Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
All patients ages 67 years and older as of December 31 of the measurement year with a history of falls, dementia or chronic kidney disease. Each of the four rates in the measure has a different denominator:
Rate 1 denominator: Patients with an accidental fall or hip fracture (Note: hip fractures are used as a proxy for identifying accidental falls). Individuals with either of the following on or between January 1 of the year prior to the measurement year and December 1 of the measurement year meet criteria:
-An accidental fall (Falls Value Set).
-An outpatient visit (Outpatient Value Set), an observation visit (Observation Value Set) or an ED visit (ED Value Set), with a hip fracture (Hip Fractures Value Set).
-An acute or nonacute inpatient discharge with a hip fracture (Hip Fractures Value Set). To identify acute and nonacute inpatient discharges: 1) Identify all acute and nonacute inpatient stays (Inpatient Stay Value Set). 2) Identify the discharge date for the stay.

Rate 2 denominator: Patients with a diagnosis of dementia (Dementia Value Set) or a dispensed dementia medication (Table DDE-C) on or between January 1 of the year prior to the measurement year and December 1 of the measurement year.

Rate 3 denominator: Patients with chronic kidney disease as identified by a diagnosis of ESRD (ESRD Value Set), stage 4 chronic kidney disease (CKD Stage 4 Value Set) or kidney transplant (Kidney Transplant Value Set) on or between January 1 of the year prior to the measurement year and December 1 of the measurement year.

Rate 4 denominator: The sum of the denominators for rates 1, 2 and 3

Note: Patients with more than one disease or condition may appear in the measure multiple times (i.e., in each indicator for which they qualify).

See S.2.b for all Value Sets

Table DDE-C: Prescriptions to Identify Members with Dementia
Cholinesterase inhibitors:
Donepezil, Galantamine, Rivastigmine

Miscellaneous central nervous system agents:
Memantine

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
The following are exclusions for the condition-specific rates and total rate:

For those who meet denominator criteria for the history of falls rate (Rate 1): exclude those with a diagnosis of psychosis, schizophrenia, bipolar disorder or seizure disorder.

For those who meet denominator criteria for those with dementia rate (Rate 2): exclude those with a diagnosis of psychosis, schizophrenia or bipolar disorder.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
For those who meet denominator criteria for the history of falls rate (Rate 1): Exclude patients with a diagnosis of psychosis (Psychosis Value Set), schizophrenia (Schizophrenia Value Set), bipolar disorder (Bipolar Disorder Value Set; Other Bipolar Disorder Value Set) or seizure disorder (Seizure Disorders Value Set) on or between January 1 of the year prior to the measurement year and December 1 of the measurement year.

For those who meet denominator criteria for those with dementia rate (Rate 2): Exclude patients with a diagnosis of psychosis (Psychosis Value Set), schizophrenia (Schizophrenia Value Set) or bipolar disorder (Bipolar Disorder Value Set; Other Bipolar Disorder Value Set) on or between January 1 of the year prior to the measurement year and December 1 of the measurement year.

See S.2.b for all Value Sets

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
No risk adjustment or risk stratification

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
N/A

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
N/A

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **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.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
Step 1. Determine the eligible population: All patients 67 years of age and older as of the end (i.e., December 31) of the measurement year.

Step 2: Identify the denominators for each of the four rates:
Rate 1: Those in the eligible population with a history of falls (see S.9 for details) on or between January 1 of the year prior to the measurement year and December 1 of the measurement year. Exclude patients with a diagnosis of psychosis, schizophrenia, bipolar disorder, or seizure disorder (see S.11 for details). Identify the index episode start date.
Rate 2: Those in the eligible population with a dementia (see S.9 for details) on or between January 1 of the year prior to the measurement year and December 1 of the measurement year. Exclude patients with a diagnosis of psychosis, schizophrenia or bipolar disorder (see S.11 for details). Identify the index episode start date.
Rate 3: Those in the eligible population with end stage renal disease (see S.9 for details) on or between January 1 of the year prior to the measurement year and December 1 of the measurement year. Identify the index episode start date.
Rate 4: The sum of denominators for Rates 1, 2 and 3.

Step 3: Identify the numerators: Individuals in each of the denominators who have received at least one potentially harmful medication on or after the index episode start date (see definitions of potentially harmful medications for each numerator in section S.6).

Step 4: Calculate the rates:
Rate 1 – Numerator 1 divided by denominator 1.
Rate 2 – Numerator 2 divided by denominator 2.

Rate 3 – Numerator 3 divided by denominator 3.
Rate 4 – The sum of the three numerators divided by the sum of the three denominators.

Note: for this measure a lower rate indicates better performance for all four rates.

Index Episode Start Date. The earliest diagnosis, procedure or prescription between January 1 of the year prior to the measurement year and December 1 of the measurement year.
For an outpatient claim/encounter, the IESD is the date of service.
For an inpatient claim/encounter, the IESD is the discharge date.
For dispensed prescriptions, the IESD is the dispense date.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
N/A

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
N/A

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims, Electronic Clinical Data, Electronic Clinical Data : Pharmacy

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
This measure is based on administrative claims collected in the course of providing care to health plan members. NCQA collects the Healthcare Effectiveness Data and Information Set (HEDIS) data for this measure directly from Health Management Organizations and Preferred Provider Organizations via NCQA's online data submission system.

S.25. **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.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Health Plan, Integrated Delivery System

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic, Pharmacy
If other:

S.28. **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

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
DDE_Testing_Final.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Potentially Harmful Drug-Disease Interactions in the Elderly
Date of Submission: Click here to enter a date
Type of Measure:

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

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient

preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who receive smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ administrative claims	☒ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

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

N/A

1.3. What are the dates of the data used in testing? 2014

HEDIS submission data from 2014

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.26***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☒ health plan	☒ health plan
☐ other: Click here to describe	☐ other: Click here to describe

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

Empiric reliability and validity statistics were calculated from HEDIS data that included 412 Medicare health plans. This included all Medicare health plans submitting data to NCQA for HEDIS. The plans were geographically diverse and varied in size.

Face Validity: This measure was tested for face validity with two panels of experts. See Additional Information: Ad.1. Workgroup/Expert Panel Involved in Measure Development for names and affiliations of expert panel members.

- The Geriatric Measurement Advisory Panel (GMAP) included 11 experts in geriatrics, including representation by consumers, health plans, health care providers and policy makers.
- NCQA's Committee on Performance Measurement (CPM) oversees the evolution of the measurement set and includes representation by purchasers, consumers, health plans, health care providers and policy makers. This panel is made up of 16 members. The CPM is organized and managed by NCQA and reports to the NCQA Board of Directors and is responsible for advising NCQA staff on the development and maintenance of performance measures. CPM members reflect the diversity of constituencies that performance measurement serves; some bring other perspectives and additional expertise in quality management and the science of measurement.

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*)

Data from the HEDIS submission for 2014 are summarized at the health plan level by numerator rate. Below is a description of the sample. It includes number of health plans included that reported this measure for HEDIS and the median eligible population for the measure across health plans.

Measure	Number of Plans	Median number of eligible patients per plan
Rate 1 (History of Falls)	387	534
Rate 2 (Dementia)	385	579
Rate 3 (Chronic Kidney Disease)	356	297
Rate 4 (Total)	412	1213

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 patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).
N/A

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*)

In order to assess measure precision in the context of the observed variability across accountable entities, we utilized the reliability estimate proposed by Adams (2009). The following is quoted from the tutorial which focused on provider-level assessment: "Reliability is a key metric of the suitability of a measure for [provider] profiling because it describes how well one can confidently distinguish the performance of one physician from another. Conceptually, it is the ratio of signal to noise. The signal in this case is the proportion of the variability in measured performance that can be explained by real differences in performance. There are three main drivers of reliability: sample size, differences between physicians, and measurement error. At the physician level, sample size can be increased by increasing the number of patients in the physician's data as well as increasing the number of measures per patient." This approach is also relevant to health plans and other accountable entities.

Adams' approach uses a Beta-binomial model to estimate reliability; this model provides a better fit when estimating the reliability of simple pass/fail rate measures as is the case with most HEDIS® measures. The beta-binomial approach accounts for the non-normal distribution of performance within and across accountable entities. Reliability scores vary from 0.0 to 1.0. A score of zero implies that all variation is attributed to measurement error (noise or the individual accountable entity variance) whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance (across accountable entities).

Adams, J. L. The Reliability of Provider Profiling: A Tutorial. Santa Monica, California: RAND Corporation. TR-653-NCQA, 2009

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

Using 2014 Health Plan performance data, reliability for the rates in this measure are as shown below. Strong reliability is demonstrated since majority of variances is due to signal and not to noise

Rate	Beta Binomial Rate
Rate 1 (History of Falls)	0.96565
Rate 2 (Dementia)	0.97552
Rate 3 (Chronic Kidney Disease)	0.95273
Rate 4 (Total)	0.98571

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?*)

Reliability scores can vary from 0.0 to 1.0. A score of zero implies that all variation is attributed to measurement error (noise) whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance (signal). Generally, a minimum reliability score of 0.7 is used to indicate sufficient signal strength to discriminate performance between accountable entities. The testing suggests that all indicators within this measure have great reliability that is above 0.95.

2b2. VALIDITY TESTING

2b2.1. 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 <u>performance measure score</u> 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*)

2b2.2. 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)*

Method of Assessing Face Validity: NCQA has identified and refined measure management into a standardized process called the HEDIS measure life cycle.

STEP 1: NCQA staff identify areas of interest or gaps in care with clinical expert panel input. Once topics are identified, a literature review and stakeholder interviews are conducted to evaluate the importance, scientific soundness and feasibility of potential measure concepts. This information is reviewed by NCQA's Measurement Advisory Panels (MAPs), the Technical Measurement Advisory Panel (TMAP) and the Committee on Performance Measurement (CPM) as well as other panels as necessary.

STEP 2: MAPs participate in the development and testing of measures by advising on measure specification, testing plans and testing results demonstrating reliability, validity and feasibility of potential measures. The CPM uses testing results and proposed final specifications to determine if the measure will move forward to Public Comment.

STEP 3: Public Comment is a 30-day period of review that allows interested parties to offer feedback to NCQA and the CPM about new measures or about changes to existing measures. MAPs consider all comments and advise NCQA staff on appropriate recommendations brought to the CPM. The CPM reviews all comments before making a final decision about Public Comment measures. New measures and changes to existing measures approved by the CPM and NCQAs Board of Directors will be included in the next HEDIS year and reported as first-year measures.

STEP 4: All new measures are collected, but results are not publicly reported in the first year. This period guarantees that a measure can be effectively collected, reported and audited before it is used for public accountability or accreditation. NCQA's experience is that the first year of large-scale data collection often reveals unanticipated issues. NCQA conducts a detailed evaluation of all first-year data. The CPM uses these evaluation results to decide whether the measure should become publicly reportable or whether it needs further modifications.

STEP 5: If the measure is approved by the CPM, it will be publically reported and may be used for scoring in accreditation.

Step 6: Evaluation is the ongoing review of a measure's performance and recommendations for its modification or retirement. Each year, NCQA prioritizes measures for re-evaluation based on changes in clinical guidelines and feedback from measure users, auditors or other stakeholders. If necessary, the measure is re-evaluated for importance, scientific soundness, and feasibility with input from the MAPs. Specifications may be updated or the measure may be recommended for retirement. The CPM reviews recommendations from the evaluation process and approves or rejects the recommendation. If approved, the change is included in the next year's HEDIS Volume.

Method of Testing Construct Validity: We empirically tested for construct validity by exploring whether the indicators within this measure were correlated with each other and with another measure of medication safety. We hypothesized that organizations that perform well on one of the three indicators should perform well on the other indicators as well as the other medication safety measure. To test these correlations we used a Pearson correlation test. This test estimates the strength of the linear association between two continuous variables; the magnitude of correlation ranges from -1 and +1. A value of 1 indicates a perfect linear dependence in which increasing values on one variable is associated with increasing values of the second variable. A value of 0 indicates no linear association. A value of -1 indicates a perfect linear relationship in which increasing values of the first variable is associated with decreasing values of the second variable.

Method for ICD-10 Conversion: Goal was to convert this measure to a new code set, fully consistent with the intent of the original measure.

Steps in ICD-9 to ICD-10 Conversion Process

1. NCQA staff identify ICD-10 codes to be considered based on ICD-9 codes currently in measure. Use GEM to identify ICD-10 codes that map to ICD-9 codes. Review GEM mapping in both directions (ICD-9 to ICD-10 and ICD-10 to ICD-9) to identify potential trending issues.
2. NCQA staff identify additional codes (not identified by GEM mapping step) that should be considered. Using ICD-10 tabular list and ICD-10 Index, search by diagnosis or procedure name for appropriate codes.
3. NCQA HEDIS Expert Coding Panel review NCQA staff recommendations and provide feedback.
4. As needed, NCQA Measurement Advisory Panels perform clinical review. Due to increased specificity in ICD-10, new codes and definitions require review to confirm the diagnosis or procedure is intended to be included in the scope of the measure. Not all ICD-10 recommendations are reviewed by NCQA MAP; MAP review items are identified during staff conversion or by HEDIS Expert Coding Panel.
5. Post ICD-10 code recommendations for public review and comment.
6. Reconcile public comments. Obtain additional feedback from HEDIS Expert Coding Panel and MAPs as needed.
7. NCQA staff finalize ICD-10 code recommendations.

Tools Used to Identify/Map to ICD-10
All tools used for mapping/code identification from CMS ICD-10 website (http://www.cms.gov/Medicare/Coding/ICD10/2012-ICD-10-CM-and-GEMs.html).
GEM, ICD-10 Guidelines, ICD-10-CM Tabular List of Diseases and Injuries, ICD-10-PCS Tabular List.

Expert Participation
NCQA's Geriatric Measurement Advisory Panel and Committee for Performance Measurement reviewed and provided feedback on staff recommendations. Names and credentials of the experts who served on these panels are listed under Additional Information, Ad. 1. Workgroup/Expert Panel Involved in Measure Development.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Results of Face Validity Assessment:
Step 1: This measure was developed to address potentially harmful drug-disease interactions in the elderly. NCQA and the GMAP worked together to assess conditions and medications based on the AGS Beers Criteria.

Step 2: The measure was field-tested from 2004-2005. After reviewing field test results the CPM recommended to send the measure to public comment with a majority vote in 2006.

Step 3: The measure was released for Public Comment in 2006 prior to publication in HEDIS. The CPM recommended moving this measure to first year data collection by a majority vote.

Step 4: The measure was introduced in HEDIS 2007. Organizations reported the measures in the first year and the results were analyzed for public reporting in the following year. The CPM recommended moving this measure public reporting with a majority vote.

Step 5: The measure is currently undergoing re-evaluation.

Conclusion: The measure was deemed to have the desirable attributes of a HEDIS measure in 2006 (relevance, scientific soundness, and feasibility).

Results of Construct Validity Testing: The results in Table 1a indicate that there was a moderate or high correlation between all rates with the exception of Rate 1 (History of Falls) and Rate 3 (Chronic Kidney Disease).

Table 1a. Correlations among all rates in DDE measure and the DAE measure[1]

Measure	Pearson Correlation Coefficients				
	DDE: Rate 1 (History of Falls)	DDE: Rate 2 (Dementia)	DDE: Rate 3 (Chronic Kidney Disease)	DDE: Rate 4 (Total)	DAE: High-Risk Med, 65+
DDE Rate 1 (History of Falls)					
DDE: Rate 2 (Dementia)	0.694				
DDE: Rate 3 (Chronic Kidney Disease)	0.155	0.585			
DDE: Rate 4 (Total)	0.842	0.921	0.480		
DAE: High-Risk Medication in those 65 and older	0.307	0.454	0.367	0.386	

Note: All correlations are significant at p<.05

[1]The DAE measure assesses the percentage of patients 65 and older who receive at least one high-risk medication (as defined by Table 2 of the American Geriatrics Society Beers Criteria. There is no disease/condition requirement for this measure.

ICD-10 Conversion:

Summary of Stakeholder Comments Received

NCQA posted ICD-10 codes for public review and comment in March 2011 and March 2012. NCQA received comments from four organizations:

- Support recommendations.
- Questions about select codes.
- Recommended additional codes for consideration.

2b2.4. 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?*)

Interpretation of systematic assessment of face validity: These results indicate the MAPs and CPM showed agreement that the measures as specified will accurately differentiate quality across health plans. Our interpretation of these results is that this measure has sufficient face validity.

Interpretation of construct validity testing: Coefficients with absolute value of less than 0.3 are generally considered indicative of weak associations whereas absolute values of 0.3 or higher denote moderate to strong associations. The significance of a correlation coefficient is evaluated by testing the hypothesis that an observed coefficient calculated for the sample is different from zero. The resulting p-value indicates the probability of obtaining a difference at least as large as the one observed due to chance alone. We used a threshold of 0.05 to evaluate the test results. P-values less than this threshold imply that it is unlikely that a non-zero coefficient was observed due to chance alone. The results confirmed the hypothesis that rates in the measure are correlated with each other as well as with another measure of medication safety, suggesting they represent the same underlying quality construct of prescribing inappropriate medications for patients with the corresponding illnesses. These results indicate the measure is a valid measure of a plan's quality at managing potentially harmful drug-disease interactions.

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

2b3.1. 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 exclusions for this measure are based on the conditions specified in the individual rates that could merit the use of a medication that is listed in the Beers Criteria. While the diagnosis codes used to identify the exclusions have not been tested in the context of this measure for validity, they are widely used across practitioners and considered to be valid. HEDIS data can be used to identify the overall rate and distribution of exclusions across health plans that report HEDIS.

2b3.2. 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*)

The following table includes the rate and distribution of exclusions across health plans reporting 2014 HEDIS data.

	Number of plans	Average rate of exclusions	Standard deviation	Min	25th	50th	75th	Max
Rate 1: History of Falls[1]	384	15.3	11.5	0.0	0.0	16.8	20.9	56.0
Rate 2: Dementia[2]	382	18.9	12.8	0.0	9.1	20.6	28.4	50.9
Rate 4: Total	409	14.8	10.7	0.0	6.9	16.0	20.67	61.0

[1]For the History of Falls rate, those with a diagnosis of psychosis, schizophrenia, bipolar disorder or seizure disorder are excluded.
[2]For the Dementia rate, those with a diagnosis of psychosis, schizophrenia or bipolar disorder are excluded.

2b3.3. 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 exclusions in this measure are identified using administrative claims codes and therefore do not add much burden to collection. The results indicate that on average 15.3 percent of patients meet the exclusion criteria for the History of Falls rate and 18.9 percent meet the exclusion criteria for the Dementia rate.

2b4. 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 2b5.

2b4.1. What method of controlling for differences in case mix is used?
☐ **No risk adjustment or stratification**
☐ **Statistical risk model with** Click here to enter number of factors **risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use 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.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

2b4.4a. What were the statistical results of the analyses used to select risk factors?

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

2b4.5. 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 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

2b4.10. 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*)

2b4.11. 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*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

To demonstrate meaningful differences in performance, NCQA calculates an inter-quartile range (IQR) for each indicator. The IQR provides a measure of the dispersion of performance. The IQR can be interpreted as the difference between the 25th and 75th percentile on a measure. To determine if this difference is statistically significant, NCQA calculates an independent sample t-test of the performance difference between two randomly selected plans at the 25th and 75th percentile. The t-test method calculates a testing statistic based on the sample size, performance rate, and standardized

error of each plan. The test statistic is then compared against a normal distribution. If the p value of the test statistic is less than .05, then the two plans' performance is significantly different from each other. Using this method, we compared the performance rates of two randomly selected plans, one plan in the 25th percentile and another plan in the 75th percentile of performance. We used data from the most recent HEDIS submissions in 2014.

2b5.2. 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*)

Variation in Performance across Plans (HEDIS Results using 2014 Data)

Rate	Number of Plans	Mean	SD	10th (Better)	25th	50th	75th	90th (Worse)	IQR	P-Value
Rate 1 (History of Falls)	387	48.0%	8.3	38.8%	43.1%	47.7%	51.9%	58.5%	8.8	0.0026
Rate 2 (Dementia)	385	48.5%	9.1	39.2%	42.8%	46.8%	52.8%	61.0%	10	0
Rate 3 (Chronic Kidney Disease)	356	9.6%	6.1	3.9%	5.8%	8.1%	12.0%	17.1%	6.2	0
Rate 4 (Total)	412	41.5%	7.9	33.5%	36.7%	40.3%	44.9%	51.3%	8.2	<0.001

IQR: Interquartile range
p-value: P-value of independent samples t-test comparing plans at the 25th percentile to plans at the 75th percentile.

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

The results above indicate there is a 6-10% gap in performance between the 25th and 75th performing plans. Plans at the 25th and 75th percentile have a statistically significant difference in performance. The largest gap in performance is for the dementia rate which had a 10% gap in performance between plans at the 25th and 75th percentiles.

2b6. 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 SDS 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 SDS 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.***

2b6.1. 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)

2b6.2. 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*)

2b6.3. 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*)

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

Plans collect this measure using all administrative data sources. NCQA's audit process checks that plans' measure calculations are not biased due to missing data.

2b7.2. 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*)

Plans collect this measure using all administrative data sources. NCQA's audit process checks that plans' measure calculations are not biased due to missing data.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

Plans collect this measure using all administrative data sources. NCQA's audit process checks that plans' measure calculations are not biased due to missing data.

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

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)*

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.**

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.**

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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

NCQA recognizes that, despite the clear specifications defined for HEDIS measures, data collection and calculation methods may vary, and other errors may taint the results, diminishing the usefulness of HEDIS data for managed care organization (MCO) comparison. In order for HEDIS to reach its full potential, NCQA conducts an independent audit of all HEDIS collection and reporting processes, as well as an audit of the data which are manipulated by those processes, in order to verify that HEDIS specifications are met. NCQA has developed a precise, standardized methodology for verifying the integrity of HEDIS collection and calculation processes through a two-part program consisting of an overall information systems capabilities assessment followed by an evaluation of the MCO´s ability to comply with HEDIS specifications. NCQA-certified auditors using standard audit methodologies will help enable purchasers to make more reliable "apples-to-apples" comparisons between health plans.
The HEDIS Compliance Audit addresses the following functions:
1) information practices and control procedures
2) sampling methods and procedures
3) data integrity
4) compliance with HEDIS specifications
5) analytic file production

6) reporting and documentation

In addition to the HEDIS Audit, NCQA provides a system to allow "real-time" feedback from measure users. Our Policy Clarification Support System receives thousands of inquiries each year on over 100 measures. Through this system NCQA responds immediately to questions and identifies possible errors or inconsistencies in the implementation of the measure. This system is vital to the regular re-evaluation of NCQA measures.

Input from NCQA auditing and the Policy Clarification Support System informs the annual updating of all HEDIS measures including updating value sets and clarifying the specifications. Measures are re-evaluated on a periodic basis and when there is a significant change in evidence. During re-evaluation information from NCQA auditing and Policy Clarification Support System is used to inform evaluation of the scientific soundness and feasibility of the measure.

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)*.
Broad public use and dissemination of these measures is encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Use by health care physicians in connection with their own practices is not commercial use. Commercial use of a measure requires the prior written consent of NCQA. As used herein, "commercial use" refers to any sale, license or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. **Current and Planned Use**

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

Planned	Current Use (for current use provide URL)
	Public Reporting Health Plan Rating http://www.ncqa.org/ReportCards/HealthPlans/HealthInsurancePlanRankings/HealthPlanRatingsPreview.aspx Annual State of Health Care Quality http://www.ncqa.org/tabid/836/Default.aspx Regulatory and Accreditation Programs HEDIS®-Health Plan http://www.ncqa.org/Programs/Accreditation/HealthPlanHP.aspx HEDIS®-ACO http://www.ncqa.org/Programs/Accreditation/AccountableCareOrganizationACO.aspx HEDIS®-Physician http://www.ncqa.org/Programs/Certification/PhysicianandHospitalQualityPHQ.aspx Quality Improvement with Benchmarking (external benchmarking to multiple

	organizations) Annual State of Health Care Quality http://www.ncqa.org/tabid/836/Default.aspx

4a.1. For each CURRENT use, checked above, provide:

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

HEALTH PLAN RATINGS/REPORT CARDS: This measure is used to calculate health plan ratings which are reported in Consumer Reports and on the NCQA website. These ratings are based on performance on HEDIS measures among other factors. In 2012, a total of 455 Medicare Advantage health plans, 404 commercial health plans and 136 Medicaid health plans across 50 states were included in the ratings. In 2015 NCQA announced a change in methodology and changed Health Plan Rankings to Health Plan Ratings.
STATE OF HEALTH CARE ANNUAL REPORT: This measure is publically reported nationally and by geographic regions in the NCQA State of Health Care annual report. This annual report published by NCQA summarizes findings on quality of care. In 2012 the report included measures on 11.5 million Medicare Advantage beneficiaries in 455 Medicare Advantage health plans, 99.4 million members in 404 commercial health plans, and 14.3 million Medicaid beneficiaries in 136 plans across 50 states.
HEALTH PLAN ACCREDITATION: This measure is used in scoring for accreditation of Medicare Advantage Health Plans. In 2012, a total of 170 Medicare Advantage health plans were accredited using this measure among others covering 7.1 million Medicare beneficiaries. [REPLACE or ADD as appropriate, 336 commercial health plans covering 87 million lives; 77 Medicaid health plans covering 9.1 million lives.] Health plans are scored based on performance compared to benchmarks.
HEDIS ACCOUNTABLE CARE ORGANIZATION ACCREDITATION: This measure is used in NCQA's ACO Accreditation program, that helps health care organizations demonstrate their ability to improve quality, reduce costs and coordinate patient care. ACO standards and guidelines incorporate whole-person care coordination throughout the health care system.
HEDIS PHYSICIAN ACCREDITATION: This measure is used in NCQA's Physician Accreditation program, that helps physicians demonstrate their ability to improve quality, reduce costs and coordinate patient care.

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

4a.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.*)
N/A

4b. 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.

4b.1. Progress on Improvement. (Not required for initial endorsement unless available.)
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

Health plan performance rates for the measure have shown slight decreases (i.e., improvement) over the last two years. For 2013, 48.5 percent of individuals with a history of falls received at least one high-risk medication and in 2014 this dropped to 48.0 percent. Among individuals with dementia, 49.7 percent received at least one high-risk medication in 2013 and this dropped to 48.5 percent in 2014. Among those with chronic kidney disease, 10.4 percent received at least one high-risk medication in 2013 and this dropped to 9.6 percent in 2014. Overall, rates from 2013 to 2014 showed a slight decrease, yet the 2014 rates still suggest significant room for improvement, particularly for the history of falls and dementia rates. For all rates there is a sizeable gap between the plans at the 10th percentile and 90th percentile, demonstrating a gap in care between the best and worst performing health plans. See section 1b.2 for a summary of recent performance data from health plans.

Due to recent updates to the medications included in this measure, future rates may show greater room for improvement and variation in performance.

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
N/A

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
There were no identified unintended consequences for this measure during testing or since implementation. If this measure were to be implemented poorly, there is concern that it could lead to reduced access to medications. There will always be individual cases that will warrant the use of a potentially harmful medication. For example, antidepressants are listed as potentially harmful to patients at risk for falls, however, clinicians should weigh the relative risk of increased falls against the potential benefit of the use of antidepressants for those with severe depression.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0022 : Use of High-Risk Medications in the Elderly (DAE)

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

5a. Harmonization

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 completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
This measure and NQF 0022 have a similar focus (measuring potentially inappropriate medication use in the elderly) and reporting level (health plan), however they have different target populations. This measure targets patients with a specific condition or disease that can experience adverse effects when combined with certain medications that are recommended to be avoided for that condition. NQF 0022 targets a larger population of all older adults and assesses use of high-risk medications that have been

recommended to be avoided in all older adults.

5b. Competing Measures

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

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

Appendix

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

No appendix **Attachment:**

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** National Committee for Quality Assurance
Co.2 **Point of Contact:** Bob, Rehm, nqf@ncqa.org, 202-955-1728-
Co.3 **Measure Developer if different from Measure Steward:** National Committee for Quality Assurance
Co.4 **Point of Contact:** Bob, Rehm, nqf@ncqa.org, 202-955-1728-

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.
Geriatric Measurement Advisory Panel (GMAP):
Wade Aubry, University of California, San Francisco
Arlene Bierman, Agency for Healthcare Research and Quality
Patricia Bomba, Excellus BlueCross BlueSheild
Jennie Chin Hansen, American Geriatrics Society
Joyce Dubow, Consumer Advocate
Peter Hollmann, Brown University
Adrienne Mims, Alliant Quality
Steven Phillips, Sierra Health Services, Inc.
Eric G Tangalos, Mayo Clinic
Joan Weiss, Health Resources and Services Administration
Neil Wenger, UCLA Division of General Internal Medicine and RAND

Committee on Performance Measurement (CPM):
Bruce Bagley, MD, FAAFP, Senior Advisor to the Professional Satisfaction and Practice Sustainability effort at the American Medical Association
Andrew Baskin, MD, National Medical Director, Quality & Provider Performance Measurement, Aetna
Patrick Conway, MD, MSC, Chief Medical Officer and Deputy Administrator, Centers for Medicare and Medicaid Services
Jonathan D. Darer, MD, MPH, Chief Innovation Officer, Geisinger Health System
Helen Darling, Strategic Advisor, National Business Group of Health
Rebekah Gee, MD, MPH, FACOG, Assistant Professor, LSUHSC

Foster Gesten, MD, FACP, New York State Department of Health
Marge Ginsburg, Executive Director, Center for Healthcare Decisions
David Grossman, MD, MPH, Executive Medical Director, Population and Purchaser Strategy, Group Health
Christine S. Hunter, MD (Co- Chair), Chief Medical Officer, US Office of Personnel Management
Jeffery Kelman, MMSc, MD, Chief Medical Officer, United Stated Department of Health and Human Services
Bernadette Loftus, MD, Associate Executive Director for the Mid-Atlantic States, The Permanente Medical Group
J. Brent Pawlecki, MD, MMM, Chief Health Officer, The Goodyear Tire & Rubber Company
Susan Reinhard, PhD, RN, Senior Vice President, AARP Public Policy Institute
Eric C. Schneider, MD, MSc, FACP (Co-chair), Senior Vice President, Policy and Research, The Commonwealth Fund
Marcus Thygeson, MD, MPH, Chief Health Officer, Blue Shield of California

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2007

Ad.3 **Month and Year of most recent revision:** 05, 2016

Ad.4 **What is your frequency for review/update of this measure?** Approximately every 3 years, sooner if the clinical guidelines have changed significantly.

Ad.5 **When is the next scheduled review/update for this measure?** 12, 2017

Ad.7 **Disclaimers:** These performance measures are not clinical guidelines and do not establish a standard of medical care, and have not been tested for all potential applications.
THE MEASURES AND SPECIFICATIONS ARE PROVIDED “AS IS” WITHOUT WARRANTY OF ANY KIND.

Ad.8 **Additional Information/Comments:** NCQA Notice of Use. Broad public use and dissemination of these measures is encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Use by health care physicians in connection with their own practices is not commercial use. Commercial use of a measure requires the prior written consent of NCQA. As used herein, “commercial use” refers to any sale, license or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.

These performance measures were developed and are owned by NCQA. They are not clinical guidelines and do not establish a standard of medical care. NCQA makes no representations, warranties or endorsement about the quality of any organization or physician that uses or reports performance measures, and NCQA has no liability to anyone who relies on such measures. NCQA holds a copyright in these measures and can rescind or alter these measures at any time. Users of the measures shall not have the right to alter, enhance or otherwise modify the measures, and shall not disassemble, recompile or reverse engineer the source code or object code relating to the measures. Anyone desiring to use or reproduce the measures without modification for a noncommercial purpose may do so without obtaining approval from NCQA. All commercial uses must be approved by NCQA and are subject to a license at the discretion of NCQA.

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 3000
De.2. **Measure Title:** PACE-Acquired Pressure Ulcer/Injury Prevalence Rate
Co.1.1. **Measure Steward:** CMS
De.3. **Brief Description of Measure:** Prevalence of PACE participants on the PACE organization census with pressure ulcers/injuries in a quarter, expressed as persons with 1 or more pressure ulcers/injuries divided by the number of participants on the PACE organization's census for at least one day during the quarter.

This is a rate-based measure of skin breakdown due to pressure or pressure combined with sheer. The rate will be calculated quarterly. The target population is participants on a PACE organizations census for at least one day during the quarter.
1b.1. **Developer Rationale:** The measure was developed for a unique program—CMS-funded Projects for All-Inclusive Care of the Elderly (PACE). The goal of PACE is to provide healthcare and other services to keep PACE participants in their homes. PACE participants are age 55+, Medicare or Medicaid-eligible or dually eligible, and have been determined to be nursing home eligible as defined by their state.

Pressure injury incidence rates for the PACE program are not available. The expected range for pressure injury rates would lie between the rates for nursing home residents and the rates for persons receiving home care. The incidence of pressure injuries ranges from 0.4 percent to 38 percent in acute care hospitals, from 2 percent to 24 percent in long-term care nursing facilities, and from 0 percent to 17 percent in home care settings (Cuddigan, Berlowitz, & Ayello, 2001).

Pressure ulcer/injury rates are an important safety concern in acute care and long-term care settings. There are an estimated 2.5 million pressure ulcers/injuries per year in acute care hospitals in the United States, with a cost of $9.1 billion to $11.6 billion (Reddy, Gill, & Ronchon, 2006; Shreve, Van Den Bos, Gray, Halford, Rustagi, & Ziemkiewicz, 2010; Institute for Healthcare Improvement, 2014). In addition to increasing health care resource consumption and costs, pressure ulcers also cause pain to the patient, prolong hospital stays, and place patients at risk for other adverse events (Gorecki et al., 2009; Lyder et al., 2012; National Pressure Ulcer Advisory Panel & European Pressure Ulcer Advisory Panel, 2009). The occurrence of pressure ulcers is considered a serious consequence of substandard quality of care.

The prevention of pressure ulcers/injuries has become the focus of national policy and patient safety initiatives. NQF (2008) considers HAPUs of stages III and IV ''largely preventable, grave errors'' (p. 1). On October 1, 2008, CMS stopped reimbursing hospitals for costs of treating stage III and IV HAPUs (CMS, 2007; Stone et al., 2010). Additionally, CMS is planning to implement the Hospital-Acquired Condition (HAC) Reduction Program in the near future, under which hospitals will be penalized for excess rates of HAPUs and other HACs (CMS, 2014). National health care stakeholders, including the National Quality Strategy and the CMS Partnership for Patients and HAC Reduction Program, have identified pressure ulcers as a patient safety concern.

Citation:
Center for Medicare & Medicaid Services. (2007). FY 2008 inpatient prospective payment system final rule. Retrieved from http://www.cms.gov/Newsroom/MediaReleaseDatabase/Fact-sheets/2007-Fact-sheets-items/2007-08-012.html.

Center for Medicare & Medicaid Services. (2014). Fact sheets: CMS proposals to improve quality of care during hospital inpatient stays. Retrieved August 24, 2014, from http://www.cms.gov/Newsroom/MediaReleaseDatabase/Fact-sheets/2014-Fact-sheets-items/2014-04-30-2.html.

Cuddigan J, Berlowitz DR, Ayello EA. Pressure ulcers in America: prevalence, incidence, and implications for the future. Reston VA: National Pressure Ulcer Advisory Panel; 2001.

Gorecki, C., Brown, J. M., Nelson, E. A., Briggs, M., Schoonhoven, L., Dealey, C., … Nixon, J. (2009). Impact of pressure ulcers on quality of life in older patients: A systematic review. Journal of the American Geriatrics Society. 57(7), 1175–1183.

Institute for Healthcare Improvement. (2014). Protecting 5 million lives from harm: Overview. Cambridge, MA. Retrieved September 27, 2014, from http://www.ihi.org/engage/Initiatives/completed/5MillionLivesCampaign/Pages/default.aspx.

Lyder, C. H., Wang, Y., Metersky, M., Curry, M., Kliman, R., Verzier, N. R., & Hunt, D. R. (2012). Hospital-acquired pressure ulcers: Results from the national Medicare Patient Safety Monitoring System Study. Journal of the American Geriatrics Society, 60(9), 1603–1608.

National Pressure Ulcer Advisory Panel (NPUAP)/EPUAP. (2009). Prevention and treatment of pressure ulcers: Clinical practice guideline. Washington, DC: NPUAP.

National Quality Forum. (2008). Serious reportable events. Retrieved from https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=57355.

Reddy, M., Gill, S. S., & Rochon, P. (2006). Preventing pressure ulcers: A systematic review. The Journal of the American Medical Association, 296(8), 974–984.

Shreve, J., Van Den Bos, J., Gray, T., Halford, M., Rustagi, K., & Ziemkiewicz, E. (2010). The economic measurement of medical errors. Sponsored by Society of Actuaries' Health Section. Milliman Inc.

Stone, P. W., Glied, S. A., McNair, P. D., Matthes, N., Cohen, B., Landers, T. F., & Larson, E. L. (2010). CMS changes in reimbursement for HAIs. Medical Care, 48, 433–439. doi: 10.1097/MLR. 0b013e3181d5fb3f.

S.4. **Numerator Statement:** The total number of participants enrolled during the quarter that have at least one documented PU (of any stage) acquired while a PACE participant.
S.7. **Denominator Statement:** Number of participants on a PACE organization's census during the quarter.
S.10. **Denominator Exclusions:** Exclude persons who were not on the PACE census for at least one day during the quarter. Exclude participants who lived outside their home/assisted living setting for every day of the quarter.

De.1. **Measure Type:** Outcome
S.23. **Data Source:** Electronic Clinical Data, Management Data, Paper Medical Records
S.26. **Level of Analysis:** Facility

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

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

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

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

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

Summary of evidence:

- Pressure ulcer incidence rates for the PACE program are not available. The expected range for pressure ulcer rates would lie between the rates for nursing home residents and the rates for persons receiving home care. The incidence of pressure ulcers ranges from 0.4 percent to 38 percent in acute care hospitals, from 2 percent to 24 percent in long-term care nursing facilities, and from 0 percent to 17 percent in home care settings.

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*

Preliminary rating for evidence: ☒ Pass ☐ No Pass

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

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

- The developers collected data from a sample of 50 sites which were randomly selected out of a total of 114 PACE sites. A total of 29 of these sites submitted data from January-February 2015 for the fall rate. One site was excluded.
- The developers found a mean pressure related injury rate of 1.85 among every 100 participants (n=28) and a mean of 0.81 per 100 participations for stage 3 or above. Their testing showed some evidence of variation in pressure injury rates by academic affiliation and with metropolitan status, however due to small sample size, none of the differences were statistically significant.
- The literature selected by the developer seem to indicate that there is a performance gap in pressure ulcer related injury rates.

Disparities

- In a study of the National Medicare Patient Safety Monitoring System, the researchers observed variance by patient characteristics and States across the nation. Specifically, patients that were older, nonwhite, and with chronic conditions (e.g., congestive heart failure and

cerebrovascular disease) were more likely to develop HAPU, and the highest HAPU incidence rates were observed in the Northeast and Missouri (4.6 percent and 5.9 percent, respectively).

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*

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

Comments: ** Strong evidence exists that pressure ulcers are highly impactful, preventable events which result in excessive clinical pain and suffering to the patient as well as high expense in total cost of care. There is considerable evidence that pressure ulcers are preventable with standards of care being executed.

** This is an outcome measure. There are no data on PU incidence rates for PACE, but comparisons to NF and HH rates are relevant since the populations served are similar (mostly dual eligible beneficiaries with multiple chronic conditions and who are at the nursing home level of care as determined by clinical assessment as required for eligibility in PACE).

** Yes it is identified and is indeed supported by the stated rationale

** This measure is an outcome measure to on a quarterly basis report for PACE participants the acquired pressure ulcer/injury rate.

The evidence to measure pressure ulcer injury using a prevalence approach has been universally used and is evidence based for hospitals, nursing homes and home care agencies. Numerous publications presented are present. CMS is seeking approval to now include this approach for PACE participants .

This is a tangential approach .This measure is an outcome to assess the efficacy of care in each of the PACE providers for care and prevention of pressure ulcer injury.

1b. Performance Gap

Comments: ** There is solid evidence that a performance gap and variation of care exists in other health care sites such as acute care hospitals, long-term care facilities and Home care settings. However, no current evidence exists in the PACE program.

** Data from 29 PACE sites from Jan and Feb. 2015 was used. A mean PU rate of 1.85 among every 100 participants was found, and a mean of 0.81 per 100 participants for stage 3 or higher. The National Medicare Patient Safety Monitoring System study showed variance by patient characteristics and States. Older, non-white patients with chronic conditions were more likely to develop a HAPU.

**Yes-indeed high

** This request does identify a performance gap for participants of PACE. These frail elderly 55 and older are cared for under a per capita approach by Medicare and Medicaid. The cost and expertise to care for them is challenging and underestimated. This measure begins to identify one of the key care issues and the prevalence of skin breakdowns in this population . The measure at this point does not address disparities or subgroups but future plans are to do so.

**Evidence from National Medicare Patient Safety Monitoring System does report patients are older, nonwhite with chronic conditions have the reported prevalence of pressure ulcer injury.

Criteria 2: Scientific Acceptability of Measure Properties
2a. Reliability
2a1. Reliability Specifications
2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented. **Data source(s):** Inclusion criteria for numerator: • Include participants living at home or in assisted living facilities. • Include participants with pressure injuries that developed and were identified less than 24 hours after the participant was in an emergency room, admitted to the hospital, nursing home, skilled nursing facility, hospice facility, or rehabilitation facility. Exclusion criteria for numerator: • Exclude participants who were not enrolled in a PACE Program for at least one day during the quarter. • Exclude participants who were not in their home setting for at least one day of the quarter. For each participant, exclude participants who were only: o In a nursing home facility o In a hospice facility o In hospice care at home o In skilled nursing care, or o In a rehabilitation setting • Exclude participants whose pressure ulcer/injury was acquired before they were enrolled in PACE. • Exclude participants with other kinds of skin breakdown that developed during the quarter, such as diabetic ulcers or venous ulcers. • Exclude participants whose only skin breakdown was documented as a “Kennedy Terminal Ulcer” during the quarter. Kennedy Terminal Ulcers are not acknowledged as a pressure ulcer/injury stage by NPUAP. • Exclude participants with pressure ulcer/injury that developed and were identified less than 24 hours after a participant returned home (or to an assisted living facility). Specific data collection items and responses: • Participant No. • Age (at end of month): - Age in years if 55–89 - Age greater >89 = 90+ - Unknown = 99 • Gender: - Male = 1 - Female = 2

- Unknown = 99
• Pressure Injury No.
• Month
- January = 1
- February = 2
- Etc.
• Pressure Injury Stage
- Stage I = 1
- Stage II = 2
- Stage III = 3
- Stage IV = 4
- Unstageable = 5
- Deep Tissue = 6
- Unknown = 99

Pressure Injury as defined by the National Pressure Ulcer Advisory Panel*:

A pressure injury is localized damage to the skin and/or underlying soft tissue usually over a bony prominence or related to a medical or other device. The injury can present as intact skin or an open ulcer and may be painful. The injury occurs as a result of intense and/or prolonged pressure or pressure in combination with shear. The tolerance of soft tissue for pressure and shear may also be affected by microclimate, nutrition, perfusion, co-morbidities and condition of the soft tissue.

Pressure ulcers/injuries are characterized by stage:

Stage 1 Pressure Injury: Non-blanchable erythema of intact skin
Intact skin with a localized area of non-blanchable erythema, which may appear differently in darkly pigmented skin. Presence of blanchable erythema or changes in sensation, temperature, or firmness may precede visual changes. Color changes do not include purple or maroon discoloration; these may indicate deep tissue pressure injury.

Stage 2 Pressure Injury: Partial-thickness skin loss with exposed dermis
Partial-thickness loss of skin with exposed dermis. The wound bed is viable, pink or red, moist, and may also present as an intact or ruptured serum-filled blister. Adipose (fat) is not visible and deeper tissues are not visible. Granulation tissue, slough and eschar are not present. These injuries commonly result from adverse microclimate and shear in the skin over the pelvis and shear in the heel. This stage should not be used to describe moisture associated skin damage (MASD) including incontinence associated dermatitis (IAD), intertriginous dermatitis (ITD), medical adhesive related skin injury (MARSI), or traumatic wounds (skin tears, burns, abrasions).

Stage 3 Pressure Injury: Full-thickness skin loss
Full-thickness loss of skin, in which adipose (fat) is visible in the injury and granulation tissue and epibole (rolled wound edges) are often present. Slough and/or eschar may be visible. The depth of tissue damage varies by anatomical location; areas of significant adiposity can develop deep wounds. Undermining and tunneling may occur. Fascia, muscle, tendon, ligament, cartilage and/or bone are not exposed. If slough or eschar obscures the extent of tissue loss this is an Unstageable Pressure Injury.

Stage 4 Pressure Injury: Full-thickness skin and tissue loss

Full-thickness skin and tissue loss with exposed or directly palpable fascia, muscle, tendon, ligament, cartilage or bone in the injury. Slough and/or eschar may be visible. Epibole (rolled edges), undermining and/or tunneling often occur. Depth varies by anatomical location. If slough or eschar obscures the extent of tissue loss this is an Unstageable Pressure Injury.

Unstageable Pressure Injury: Obscured full-thickness skin and tissue loss
Full-thickness skin and tissue loss in which the extent of tissue damage within the injury cannot be confirmed because it is obscured by slough or eschar. If slough or eschar is removed, a Stage 3 or Stage 4 pressure injury will be revealed. Stable eschar (i.e. dry, adherent, intact without erythema or fluctuance) on an ischemic limb or the heel(s) should not be removed.

Deep Tissue Pressure Injury: Persistent non-blanchable deep red, maroon or purple discoloration
Intact or non-intact skin with localized area of persistent non-blanchable deep red, maroon, purple discoloration or epidermal separation revealing a dark wound bed or blood filled blister. Pain and temperature change often precede skin color changes. Discoloration may appear differently in darkly pigmented skin. This injury results from intense and/or prolonged pressure and shear forces at the bone-muscle interface. The wound may evolve rapidly to reveal the actual extent of tissue injury, or may resolve without tissue loss. If necrotic tissue, subcutaneous tissue, granulation tissue, fascia, muscle or other underlying structures are visible, this indicates a full thickness pressure injury (Unstageable, Stage 3 or Stage 4). Do not use DTPI to describe vascular, traumatic, neuropathic, or dermatologic conditions.

* This PU/I data collection will follow the NPUAP pressure ulcer/injury definition and staging categories. More information can be found in this link: http://www.npuap.org/national-pressure-ulcer-advisory-panel-npuap-announces-a-change-in-terminology-from-pressure-ulcer-to-pressure-injury-and-updates-the-stages-of-pressure-injury/

Denominator:

Number of participants on the PACE site census at least one day during the quarter.

Questions for the Committee :

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

N/A

SUMMARY OF TESTING

Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

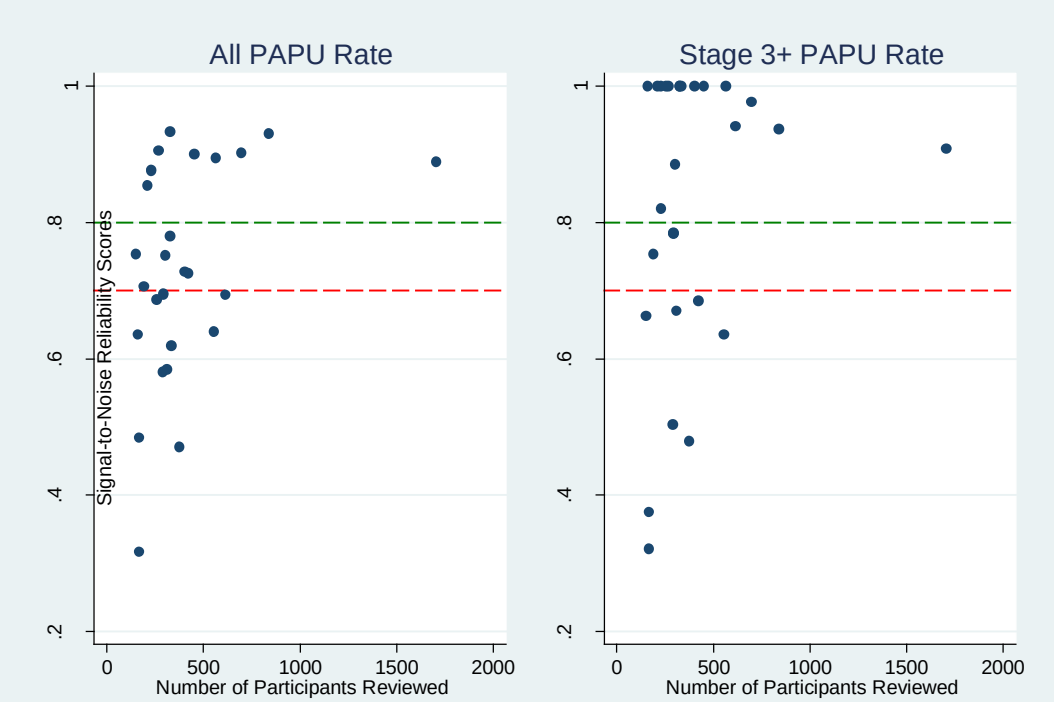
Method(s) of reliability testing Signal-to-noise analysis approach

Results of reliability testing

Table 2. Signal-to-Noise Assessment of Reliability of Pressure Ulcer/Injury Rates

	Reliability Scores	
Measures	**Mean (SD)**	**Median (Min, Max)**
PAPU/I rate (*n*=28)	0.73 (0.16)	0.73 (0.32, 0.93)
PAPU/I stage 3+ rate (*n*=28)	0.83 (0.21)	0.92 (0.33, 1.00)

Figure 1. Signal-to-Noise Reliability Assessment of PAPU/I Rates



Red line at 0.7: the acceptable reliability score; green line at 0.8: high reliability score

Guidance from the Reliability Algorithm Based on the data provided, there was empirical reliability testing using statistical tests with the measure as specified. Based on the reliability scores as "acceptable" 0.73 for all PAPU/I rates and high for PAPU/I stage 3+ rates at 0.83, reliability should be "MODERATE".

Questions for the Committee:

- *Specific questions on the method and results of reliability testing.*
- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

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

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

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, summarize the validity testing from the prior review:

N/A

SUMMARY OF TESTING

Validity testing level ☐ **Measure score** ☐ **Data element testing against a gold standard** ☒ **Both**

Method of validity testing of the measure score:

- ☐ **Face validity only**
- ☐ **Empirical validity testing of the measure score**

Validity testing method: Content validity was assessed using a national panel of pressure ulcer/injury experts to quantify experts' assessments of the validity of the PAPI numerator, denominator, and calculated rate. Content validity of the measure was analyzed by calculating item-level content validity indices (I-CVIs). The I-CVI indicates the proportion of experts who consider the item as content valid. Experts rated each component's content/face validity using a 4-point scale: 1 = very low (major modification needed), 2 = low (some modification needed), 3 = high (no modification needed but could be improved with minor changes), and 4 = very high (no modification needed). I-CVI is computed for each item by counting the number of experts giving a rating of 3 or 4 and dividing the number by the total number of experts (Polit, Beck, & Owen, 2007).

Validity testing results:

Table 3: Content Validity Results for Data Elements in the PAPU/I Data Collection Instructions

Data Element	I-CVI
PAPU/I Prevalence Rate distinguishes good from poor quality of care	0.75 (6/8)
The measure captures what this measure intends to measure:	
PAPU/I Rate	0.88 (7/8)
PAPU/I Numerator	0.88 (7/8)
PAPU/I Denominator	0.88 (7/8)
Exclusions from both Numerator and Denominator:	
Each day of quarter participant not enrolled	0.88 (7/8)
Each day of quarter participant not in home setting	
• Hospitalized more than 23 hours	1.00 (8/8)
• In emergency room more than 23 hours	0.88 (7/8)
• In a nursing home facility	1.00 (7/7)
• In a hospice facility	0.88 (7/8)
• In hospice care at home	1.00 (8/8)
• In skilled nursing care	1.00 (8/8)
• In a rehabilitation setting	1.00 (8/8)
Exclusion Criteria for Numerator:	
Pressure ulcer/injury acquired before PACE enrollment	1.00 (8/8)
Other kinds of skin breakdown that developed during the quarter (e.g. diabetic ulcers, venous ulcers)	0.75 (6/8)
Kennedy Terminal Ulcers	0.63 (5/8)
Pressure ulcers/injuries that developed and were identified less than 24 hours after a participant returned home (or to an assisted living)	0.86 (6/7)
Exclusion Criteria for Denominator:	
Deceased participants after the date of death	1.00 (7/7)
Inclusion Criteria for both Numerator and Denominator:	
Participants in assisted living facilities	1.00 (8/8)
Inclusion Criteria for Numerator:	
Pressure ulcers/injuries that developed and identified less than 24 hours after the participant was in emergency room, admitted to the hospital, nursing home, skilled nursing facility, hospice facility, or rehabilitation facility	0.83 (5/6)
Inclusion Criteria for Denominator:	
Each day a participant was on the participant census after enrolling	1.00 (7/7)

The developer commented: "A total of 8 academic experts completed content validity testing. As shown in Table 2 above, the majority of items on the content validity testing survey had good validity as indicated by an I-CVI of greater than 0.78 (16 of 20 items or 75%). In addition, none of the items was disagreed upon by 6 or more experts."

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

Content Validity of the Exclusions is described in Table 2 above.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☐ **None** ☐ **Statistical model** ☒ **Stratification**

There is stratification by 2 risk categories (Age and Gender)

Conceptual rationale for SDS factors included ? ☐ **Yes** ☒ **No**

SDS factors included in risk model? ☐ **Yes** ☒ **No**

Risk adjustment summary The developers state that it will be risk-stratified by Age and Gender given both have correlations with the performance measure. Stratification will be based on PACE site characteristics.

Questions for the Committee:

- *Is sufficient information given about the risk-adjustment methodology to assess its validity?*
- *Is an appropriate risk-adjustment strategy included in the measure?*
- *Do you agree with the developer's decision, based on their analysis, to not include SDS factors in their risk-adjustment model?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

The developer provided the following statement:

"Due to our small sample size, we did not conduct statistical analyses to determine differences in performance across PACE sites. However, the descriptive statistics indicate that there are differences in PAPU/I rates per 100 participants across PACE sites (mean = 3.67, SD = 32.25, median = 3.67, range = 0-100). After implementation, we will conduct further analyses to determine significant differences in PAPU/I rates across PACE sites."

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

N/A

2b7. Missing Data

The developer stated: "We had no missing data on the numerator or denominator for pressure ulcers from sites included in the sample."

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

Committee pre-evaluation comments

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

2a1. & 2b1. Specifications

Comments: **The elements of the measure appear to be clearly defined and follow similar conventions with other pressure ulcer classifications. Clarification needed around data source and coding specifications for administrative portion of the data collection as well as clarification on source of data for clinical parameters of the pressure ulcer.
**Specifications are clear and consistent with the evidence.
**The specifications for data collection are well established in other settings and have been proven to be moderately reliable. The inclusion and exclusion criteria for the numerator are reasonable but accuracy at each PACE program will be a challenge. There is no inconsistencies with the data as based on other similar providers to this population under CMS. Staging of pressure ulcers was based on the National Pressure Ulcer Advisory Panel, April 2016 international white paper.

2a2. Reliability Testing
Comments: **Initial reliability testing was performed utilizing the Signal-to-Noise technique and resulted in acceptable reliability of 0.73 for all PAPU rates and 0.83 for the Stage 3+ PAPU rate. Review of larger sample size and longer period of time would be helpful to reassess the reliability of this measure.
**Measure score reliability testing was done using signal-to-noise analysis. The mean PAPU incidence rate was 0.73 == acceptable for reliability. For stage 3 or higher, the mean incidence rate was 0.83 = high for reliability.
**Moderate
**The level of testing was paper and electronic data on number of pressure ulcer injuries using a prevalence approach that is reported quarterly. Individual patient data not applicable. Reliability was based on a random sample of 50 sites out of 114 PACE sites . 29 of the sites participated and provided data for one month :January-February 2015.A slight variation was shown academic settings in metro areas but otherwise too small of a sample. Additional concern only one month of data from 28 sites instead of the quarterly data expected for the measure.

2b2. Validity Testing
Comments: **Content validity was performed with a National panel of experts for the elements of the numerator, denominator and the calculated rate. The content validity testing provides a good start towards the validity of the measure. I would prefer additional validity results with other measures with pressure ulcer rates to establish convergent and discriminant validity. Is this measure presented one year too early and needs further testing?
**Content validity was tested using a national panel of PU experts to obtain experts' assessments of the validity of the PAPI numerator, denominator and calculated rate. 16 of the 20 items tested for validity were rated greater than 75%.
**This is definitely an indicator of quality in this population- the morbidity associated with this is significant and can be prevented
**Content validity was analyzed content experts using a 4 point scale item scale of 20 questions. The majority of the items had good validity by a ICVI score of greater than 0.78(16 out of 20). Also none of the items were disagreed upon by 6 or more experts.
**Reliability tesing was done by signal to noise process. Criteria robust and results demonstrated a moderate based on a 0.73 for all pressure ulcer injuries and a 0.83 for stage 3." Concern for inter rater reliability exists " via personal phone call with Dr Art Stone , board member of National Pressure Ulcer Advisory Panel.

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias
Comments: **The exclusions seem reasonable and supported from expert opinion. If the measure will be risk-stratified by age and gender to test for socioeconomic adjustments, should we have that data prior to endorsement consideration? Initial descriptive evidence points to meaningful differences existing in the data, but it is too early for a conclusion.
**Exclusions are supported by the evidence and the TEP.
The measure will be risk-stratified by age and gender given that both have correlations with the measure.
Stratification will be based on PACE site characteristics.
Descriptive statistics show there are differences in PAPU/I rates per 100 participants across PACE sites.
**Threats to validity are clearly identified in the inclusion and exclusion criteria. Major issue is will be PACE facilities with paper data collection and adhering to the criteria

Criterion 3. Feasibility

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.

- This measure is generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score, and/or, Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)
- Some data elements are in defined fields in electronic sources
- Some PACE Organizations do not use electronic medical records. All organizations will abstract data manually for this measure from either their electronic or paper charts.
- Overall, the data collection time was reasonable, around 4 hours with less than an hour for data submission when the developer conducted a survey with PACE organizations to collect information on their experiences with data collection.
- There is a perceived data collection burden, however, this is outweighed by the usefulness of the data for quality improvement and distinguishing PACE sites based on their quality of care.
- Because of the high reported ease of obtaining the data, the developer anticipates that the perceived data collection burden will decrease as sites become more familiar with the data collection and submission process.
- No fees or licensing requirements to use any aspect of the measure as specified, were reported.

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

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **As this measure is specifically designed for one program administered by CMS, it should be feasible for its limited use in that setting given the access to the multiple forms of data needed for this measure. Further expansion in the use of this measure would be limited by the ability to capture all data elements across the administrative and clinical assessments needed.
**Many PACE sites do not have EMRs. All organizations will abstract data manually from electronic and paper records. Data collection time was approx. 4 hours Burden is outweighed by usefulness of the data for QI
**Highly feasible
**This measure is feasible but will require thoughtful ways of collecting the data expertly and reporting data based on inclusion and exclusion definitions.
The PACE participants are under reported and under estimated in the amount of care and financial struggles each PACE has,

Criterion 4: Usability and Use

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

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

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

OR

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

Accountability program details:

- This is a new measure. The developer is evaluating its use in upcoming PACE quality programs.
- The developer is considering the use of the PACE-Acquired Pressure Ulcer/Injury Prevalence Rate in accountability applications within the next two years.

Improvement results:

- Not applicable as this is a newly developed measure.

Unexpected findings (positive or negative) during implementation:

- Not applicable as this is a newly developed measure.

Potential harms:

- No negative unintended consequences have been identified.

Feedback :

- Developer did not identify any specific feedback loops related to this 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

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **The measure is not currently in use as it is newly proposed. It is specifically designed for implementation in the PACE program within the next two years.
**This measure is not currently used in public reporting or in an accountability program. CMS is considering the use of this measure in accountability applications within the next 2 years.
**yes this can further the quality of care. The unintended consequences would occur fi PU present on admission were not captured.
**This will be first time that PACE outcomes will be reported .
This beginning reporting and future plans will be to begin to cohort and identify pressure ulcer injury in PACE participants based on race, gender and co morbid conditions.

Criterion 5: Related and Competing Measures

Related or competing measures

- 0201 : Pressure ulcer prevalence (hospital acquired)
- 0538 : Pressure Ulcer Prevention and Care
- 0678 : Percent of Residents or Patients with Pressure Ulcers That Are New or Worsened (Short-Stay)
- 0679 : Percent of High Risk Residents with Pressure Ulcers (Long Stay)

Harmonization

- The measures being developed for the PACE program are not closely aligned with any of the four endorsed pressure ulcer/injury measures. It appears that they all use the same conceptual definition of a pressure ulcer/injury, although the data sources and methods differ enough from each other to result in concrete definitional differences. In addition to differences in data sources, none of the related measures collect data on pressure injuries acquired in the home setting or pressure ulcers/injuries in PACE participants.

Pre-meeting public and member comments

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: PACE-Acquired Pressure Ulcer/Injury Prevalence Rate
IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: 5/13/2016

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus
The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes
3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.
4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.
5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.
6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

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

☒ Health outcome: Pressure Ulcers

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

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

☐ Process: Click here to name the process

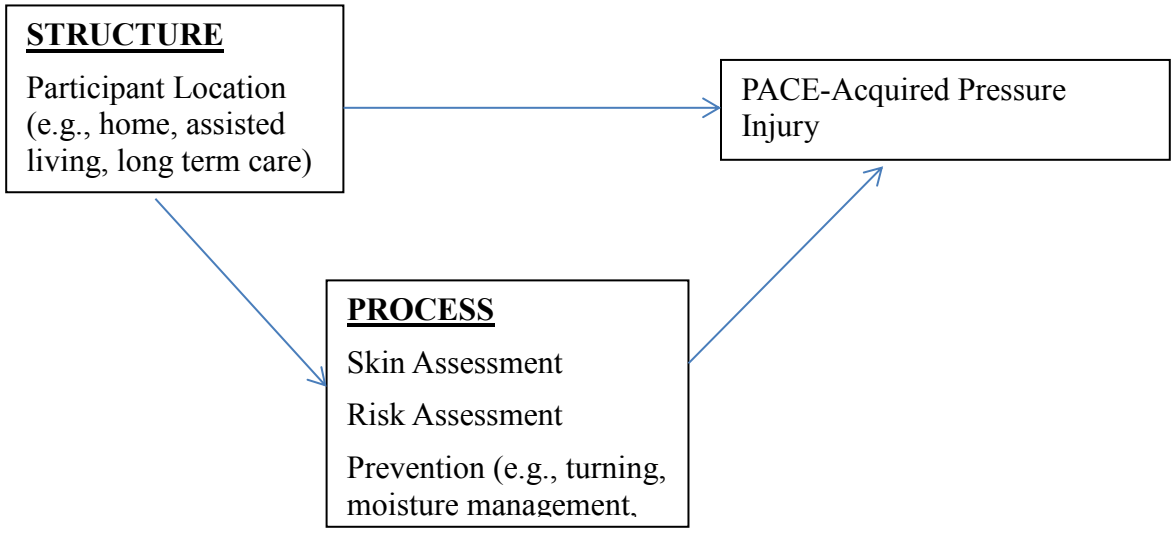
☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

_____

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.



1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

Pressure ulcers are a serious problem in the U.S. health care system, and their prevention has become a national policy issue. Several national health care improvement organizations—including the National Quality Strategy, the Partnership for Patients, and the CMS HAC Reduction Program—have identified pressure ulcers as a patient safety concern.

Reducing the occurrence of pressure ulcers is a goal of the Partnership for Patients. Pressure ulcers can cause pain and serious infections, prolong hospital stays for patients, and lead to increased health care costs.

Pressure ulcer incidence rates for the PACE program are not available. The expected range for pressure ulcer rates would lie between the rates for nursing home residents and the rates for persons receiving home care. The incidence of pressure ulcers ranges from 0.4 percent to 38 percent in acute care hospitals, from 2 percent to 24 percent in long-term care nursing facilities, and from 0 percent to 17 percent in home care settings (Cuddigan, Berlowitz, & Ayello, 2001).

Cuddigan J, Berlowitz DR, Ayello EA. Pressure ulcers in America: prevalence, incidence, and implications for the future. Reston VA: National Pressure Ulcer Advisory Panel; 2001.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

_____

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

_____

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

_____

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: Click here to enter date range

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

_____

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

NATIONAL QUALITY FORUM

Measure Information

This document contains the information submitted by measure developers/stewards, but is organized according to NQF's measure evaluation criteria and process. The item numbers refer to those in the submission form but may be in a slightly different order here. In general, the item numbers also reference the related criteria (e.g., item 1b.1 relates to subcriterion 1b).

Brief Measure Information

NQF #: 3000
De.2. **Measure Title:** PACE-Acquired Pressure Ulcer/Injury Prevalence Rate
Co.1.1. **Measure Steward:** CMS
De.3. **Brief Description of Measure:** Prevalence of PACE participants on the PACE organization census with pressure ulcers/injuries in a quarter, expressed as persons with 1 or more pressure ulcers/injuries divided by the number of participants on the PACE organization's census for at least one day during the quarter.

This is a rate-based measure of skin breakdown due to pressure or pressure combined with sheer. The rate will be calculated quarterly. The target population is participants on a PACE organizations census for at least one day during the quarter.
1b.1. **Developer Rationale:** The measure was developed for a unique program—CMS-funded Projects for All-Inclusive Care of the Elderly (PACE). The goal of PACE is to provide healthcare and other services to keep PACE participants in their homes. PACE participants are age 55+, Medicare or Medicaid-eligible or dually eligible, and have been determined to be nursing home eligible as defined by their state.

Pressure injury incidence rates for the PACE program are not available. The expected range for pressure injury rates would lie between the rates for nursing home residents and the rates for persons receiving home care. The incidence of pressure injuries ranges from 0.4 percent to 38 percent in acute care hospitals, from 2 percent to 24 percent in long-term care nursing facilities, and from 0 percent to 17 percent in home care settings (Cuddigan, Berlowitz, & Ayello, 2001).

Pressure ulcer/injury rates are an important safety concern in acute care and long-term care settings. There are an estimated 2.5 million pressure ulcers/injuries per year in acute care hospitals in the United States, with a cost of $9.1 billion to $11.6 billion (Reddy, Gill, & Ronchon, 2006; Shreve, Van Den Bos, Gray, Halford, Rustagi, & Ziemkiewicz, 2010; Institute for Healthcare Improvement, 2014). In addition to increasing health care resource consumption and costs, pressure ulcers also cause pain to the patient, prolong hospital stays, and place patients at risk for other adverse events (Gorecki et al., 2009; Lyder et al., 2012; National Pressure Ulcer Advisory Panel & European Pressure Ulcer Advisory Panel, 2009). The occurrence of pressure ulcers is considered a serious consequence of substandard quality of care.

The prevention of pressure ulcers/injuries has become the focus of national policy and patient safety initiatives. NQF (2008) considers HAPUs of stages III and IV ''largely preventable, grave errors'' (p. 1). On October 1, 2008, CMS stopped reimbursing hospitals for costs of treating stage III and IV HAPUs (CMS, 2007; Stone et al., 2010). Additionally, CMS is planning to implement the Hospital-Acquired Condition (HAC) Reduction Program in the near future, under which hospitals will be penalized for excess rates of HAPUs and other HACs (CMS, 2014). National health care stakeholders, including the National Quality Strategy and the CMS Partnership for Patients and HAC

Reduction Program, have identified pressure ulcers as a patient safety concern.

Citation:

Center for Medicare & Medicaid Services. (2007). FY 2008 inpatient prospective payment system final rule. Retrieved from http://www.cms.gov/Newsroom/MediaReleaseDatabase/Fact-sheets/2007-Fact-sheets-items/2007-08-012.html.

Center for Medicare & Medicaid Services. (2014). Fact sheets: CMS proposals to improve quality of care during hospital inpatient stays. Retrieved August 24, 2014, from http://www.cms.gov/Newsroom/MediaReleaseDatabase/Fact-sheets/2014-Fact-sheets-items/2014-04-30-2.html.

Cuddigan J, Berlowitz DR, Ayello EA. Pressure ulcers in America: prevalence, incidence, and implications for the future. Reston VA: National Pressure Ulcer Advisory Panel; 2001.

Gorecki, C., Brown, J. M., Nelson, E. A., Briggs, M., Schoonhoven, L., Dealey, C., … Nixon, J. (2009). Impact of pressure ulcers on quality of life in older patients: A systematic review. Journal of the American Geriatrics Society. 57(7), 1175–1183.

Institute for Healthcare Improvement. (2014). Protecting 5 million lives from harm: Overview. Cambridge, MA. Retrieved September 27, 2014, from http://www.ihi.org/engage/Initiatives/completed/5MillionLivesCampaign/Pages/default.aspx.

Lyder, C. H., Wang, Y., Metersky, M., Curry, M., Kliman, R., Verzier, N. R., & Hunt, D. R. (2012). Hospital-acquired pressure ulcers: Results from the national Medicare Patient Safety Monitoring System Study. Journal of the American Geriatrics Society, 60(9), 1603–1608.

National Pressure Ulcer Advisory Panel (NPUAP)/EPUAP. (2009). Prevention and treatment of pressure ulcers: Clinical practice guideline. Washington, DC: NPUAP.

National Quality Forum. (2008). Serious reportable events. Retrieved from https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=57355.

Reddy, M., Gill, S. S., & Rochon, P. (2006). Preventing pressure ulcers: A systematic review. The Journal of the American Medical Association, 296(8), 974–984.

Shreve, J., Van Den Bos, J., Gray, T., Halford, M., Rustagi, K., & Ziemkiewicz, E. (2010). The economic measurement of medical errors. Sponsored by Society of Actuaries' Health Section. Milliman Inc.

Stone, P. W., Glied, S. A., McNair, P. D., Matthes, N., Cohen, B., Landers, T. F., & Larson, E. L. (2010). CMS changes in reimbursement for HAIs. Medical Care, 48, 433–439. doi: 10.1097/MLR. 0b013e3181d5fb3f.

S.4. **Numerator Statement:** The total number of participants enrolled during the quarter that have at least one documented PU (of any stage) acquired while a PACE participant.
S.7. **Denominator Statement:** Number of participants on a PACE organization's census during the quarter.
S.10. **Denominator Exclusions:** Exclude persons who were not on the PACE census for at least one day during the quarter. Exclude participants who lived outside their home/assisted living setting for every day of the quarter.

De.1. **Measure Type:** Outcome
S.23. **Data Source:** Electronic Clinical Data, Management Data, Paper Medical Records
S.26. **Level of Analysis:** Facility

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

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

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

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

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

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

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., the benefits or improvements in quality envisioned by use of this measure)*

The measure was developed for a unique program—CMS-funded Projects for All-Inclusive Care of the Elderly (PACE). The goal of PACE is to provide healthcare and other services to keep PACE participants in their homes. PACE participants are age 55+, Medicare or Medicaid-eligible or dually eligible, and have been determined to be nursing home eligible as defined by their state.

Pressure injury incidence rates for the PACE program are not available. The expected range for pressure injury rates would lie between the rates for nursing home residents and the rates for persons receiving home care. The incidence of pressure injuries ranges from 0.4 percent to 38 percent in acute care hospitals, from 2 percent to 24 percent in long-term care nursing facilities, and from 0 percent to 17 percent in home care settings (Cuddigan, Berlowitz, & Ayello, 2001).

Pressure ulcer/injury rates are an important safety concern in acute care and long-term care settings. There are an estimated 2.5 million pressure ulcers/injuries per year in acute care hospitals in the United States, with a cost of $9.1 billion to $11.6 billion (Reddy, Gill, & Ronchon, 2006; Shreve, Van Den Bos, Gray, Halford, Rustagi, & Ziemkiewicz, 2010; Institute for Healthcare Improvement, 2014). In addition to increasing health care resource consumption and costs, pressure ulcers also cause pain to the patient, prolong hospital stays, and place patients at risk for other adverse events (Gorecki et al., 2009; Lyder et al., 2012; National Pressure Ulcer Advisory Panel & European Pressure Ulcer Advisory Panel, 2009). The occurrence of pressure ulcers is considered a serious consequence of substandard quality of care.

The prevention of pressure ulcers/injuries has become the focus of national policy and patient safety initiatives. NQF (2008) considers HAPUs of stages III and IV ‘‘largely preventable, grave errors’’ (p. 1). On October 1, 2008, CMS stopped reimbursing hospitals for costs of treating stage III and IV HAPUs (CMS, 2007; Stone et al., 2010). Additionally, CMS is planning to implement the Hospital-Acquired Condition (HAC) Reduction Program in the near

future, under which hospitals will be penalized for excess rates of HAPUs and other HACs (CMS, 2014). National health care stakeholders, including the National Quality Strategy and the CMS Partnership for Patients and HAC Reduction Program, have identified pressure ulcers as a patient safety concern.

Citation:
Center for Medicare & Medicaid Services. (2007). FY 2008 inpatient prospective payment system final rule. Retrieved from http://www.cms.gov/Newsroom/MediaReleaseDatabase/Fact-sheets/2007-Fact-sheets-items/2007-08-012.html.

Center for Medicare & Medicaid Services. (2014). Fact sheets: CMS proposals to improve quality of care during hospital inpatient stays. Retrieved August 24, 2014, from http://www.cms.gov/Newsroom/MediaReleaseDatabase/Fact-sheets/2014-Fact-sheets-items/2014-04-30-2.html.

Cuddigan J, Berlowitz DR, Ayello EA. Pressure ulcers in America: prevalence, incidence, and implications for the future. Reston VA: National Pressure Ulcer Advisory Panel; 2001.

Gorecki, C., Brown, J. M., Nelson, E. A., Briggs, M., Schoonhoven, L., Dealey, C., … Nixon, J. (2009). Impact of pressure ulcers on quality of life in older patients: A systematic review. Journal of the American Geriatrics Society. 57(7), 1175–1183.

Institute for Healthcare Improvement. (2014). Protecting 5 million lives from harm: Overview. Cambridge, MA. Retrieved September 27, 2014, from http://www.ihi.org/engage/Initiatives/completed/5MillionLivesCampaign/Pages/default.aspx.

Lyder, C. H., Wang, Y., Metersky, M., Curry, M., Kliman, R., Verzier, N. R., & Hunt, D. R. (2012). Hospital-acquired pressure ulcers: Results from the national Medicare Patient Safety Monitoring System Study. Journal of the American Geriatrics Society, 60(9), 1603–1608.

National Pressure Ulcer Advisory Panel (NPUAP)/EPUAP. (2009). Prevention and treatment of pressure ulcers: Clinical practice guideline. Washington, DC: NPUAP.

National Quality Forum. (2008). Serious reportable events. Retrieved from https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=57355.

Reddy, M., Gill, S. S., & Rochon, P. (2006). Preventing pressure ulcers: A systematic review. The Journal of the American Medical Association, 296(8), 974–984.

Shreve, J., Van Den Bos, J., Gray, T., Halford, M., Rustagi, K., & Ziemkiewicz, E. (2010). The economic measurement of medical errors. Sponsored by Society of Actuaries' Health Section. Milliman Inc.

Stone, P. W., Glied, S. A., McNair, P. D., Matthes, N., Cohen, B., Landers, T. F., & Larson, E. L. (2010). CMS changes in reimbursement for HAIs. Medical Care, 48, 433–439. doi: 10.1097/MLR. 0b013e3181d5fb3f.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

A sample of 50 organizations was randomly selected out of a total of 114 PACE organizations. Additionally, the oldest and two newest PACE organizations were included in the sample. A total of 29 of these organizations submitted data from January and February 2015 for pressure injury prevalence. One (1) of the organizations had only one (1) participant in January and February 2015 and this organization was excluded from the summary and

analysis as it is an extreme outlier that cannot provide reliable pressure injury rates. The table below shows the organization-level descriptive statistics for PACE-Acquired Pressure Ulcer/Injury (PAPU/I) rates and stage 3 or above PACE-Acquired Pressure Ulcer/Injury (PAPU/I stage 3+) rates for the 28 organizations.

Pressure Injuries-Related Measures(n) Mean, Std. Dev., Median, Minimum, Maximum
Average participants reviewed for pressure injuries (n = 28) 198.80, 153.76, 152.75, 76, 852
Number of participants with PACE-acquired pressure injuries among every 100 participants (n = 28) 1.85, 1.40, 1.44, 0.31, 5.60
Number of participants with PACE-acquired stage 3 or above pressure injuries (n = 28) 0.81, 1.06, 0.38, 0, 3.47

Testing showed some evidence of variation in pressure injury rates by academic affiliation and with metropolitan status, however due to small sample size, none of these differences were statistically significant.

Descriptive Summary of All Pressure Injury Rates by Academic Affiliation and Location (n = 28)

Affiliated with Academic Medical Center

Yes/No	N	Mean	SD	Median	z-stat*	p-value
Yes	3	2.23	1.72	1.71	-0.56	0.58
No	25	1.80	1.39	1.32		

Location	N	Mean	SD	Median	Chi-squared**	p-value
Metropolitan	22	1.91	1.24	1.71	1.52	0.47
Micropolitan	2	1.11	0.64	1.11		
Non-metropolitan	4	1.86	2.51	0.74		

* Wilcoxon-Mann-Whitney test was used to examine difference in all PAPU rates by academic affiliation.
** Kruskal Wallis test was used to examine difference in all PAPU rates by location.

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.

Pressure ulcer data have not yet been collected across PACE sites. Thus, evidence currently available is primarily from hospital- or nursing home-based studies. It is estimated that there are approximately 2.5 million pressure ulcers in acute care hospitals in the United States (Sherve, 2010), or a nationwide hospital-associated pressure ulcer (HAPU) incidence rate of 4.5 percent (Lyder et al., 2012). In another study using data from a national quality indicators database, researchers observed variance in pressure ulcers by unit types (Bergquist-Beringer, Dong, He, & Dunton, 2013). Critical care units had the highest rates (8.1 percent) relative to step-down units (3.7 percent), medical units (3.1 percent), surgical units (2.4 percent), and medical-surgical combined units (2.6 percent). The researchers also reported that the frequency of interventions to prevent pressure ulcers also varied among at-risk patients. For example, the researchers reported that only 56.3 percent of at-risk patients received nutrition support. Other studies of pressure ulcers in U.S. acute care hospitals indicated that the occurrence of pressure ulcers during hospitalization was related to hospital characteristics (e.g., bed size, teaching status, and Magnet status), nursing resources, and work conditions (Park, Boyle, Bergqust-Beringer, Staggs, & Dunton, 2014; Choi, Bergquist-Beringer, & Staggs, 2013).

Researchers have also studied pressure ulcers among nursing home residents. Park-Lee and Caffrey (2009) report that about 11 percent of the 1.5 million U.S. nursing residents in 2004 developed at least one pressure ulcer. They also found that only 35 percent of residents with stage II or higher pressure ulcers received wound care by specially trained professionals or staff.

There are performance gaps in pressure ulcers. In 2004, pressure ulcer rates in U.S. nursing homes ranged from 2 percent to 28 percent (Park-Lee & Caffrey, 2009). In 2010, a study using data obtained from acute care hospital units found that HAPU rates differed by unit type (Bergquist-Beringer, Dong, He, & Dunton, 2013).

Citations:
Bergquist-Beringer, S., Dong, L., He, J., & Dunton, N. (2013). Pressure ulcers and prevention among acute care hospitals in the United States. The Joint Commission Journal on Quality and Patient Safety, 39, 404–414.

Choi, J., Bergquist-Beringer, S., & Staggs, V. S. (2013). Linking RN workgroup job satisfaction to pressure ulcers among older adults on acute care hospital units. Research in Nursing & Health, 36(2), 181–190.

Lyder, C. H., Wang, Y., Metersky, M., Curry, M., Kliman, R., Verzier, N. R., & Hunt, D. R. (2012). Hospital-acquired pressure ulcers: Results from the national Medicare Patient Safety Monitoring System study. Journal of the American Geriatrics Society, 60(9), 1603–1608.

Park, S. H., Boyle, D. K., Bergquist-Beringer, S., Staggs, V. S., & Dunton, N. E. (2014). Concurrent and lagged effects of registered nurse turnover and staffing on unit-acquired pressure ulcers. Health Services Research, 49(4), 1205–1225.

Park-Lee, E., & Caffrey, C. (2009) Pressure ulcers among nursing home residents: United States, 2004. NCHS Data Brief, 14. Retrieved fromhttp://www.cdc.gov/nchs/data/databriefs/db14.pdf.

Shreve, J., Van Den Bos, J., Gray, T., Halford, M., Rustagi, K., & Ziemkiewicz, E. (2010). The economic measurement of medical errors. Sponsored by Society of Actuaries' Health Section. Milliman Inc.

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*
Not applicable.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
In a study of the National Medicare Patient Safety Monitoring System, the researchers observed variance by patient characteristics and States across the nation (Lyder et al., 2012). Specifically, patients that were older, nonwhite, and with chronic conditions (e.g., congestive heart failure and cerebrovascular disease) were more likely to develop HAPU, and the highest HAPU incidence rates were observed in the Northeast and Missouri (4.6 percent and 5.9 percent, respectively).

Citation:
Lyder, C. H., Wang, Y., Metersky, M., Curry, M., Kliman, R., Verzier, N. R., & Hunt, D. R. (2012). Hospital-acquired pressure ulcers: Results from the national Medicare Patient Safety Monitoring System study. Journal of the American Geriatrics Society, 60(9), 1603–1608.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF;
  OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of

patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, Patient/societal consequences of poor quality
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
Pressure ulcer/injury rates are an important safety concern in acute care and long-term care settings. There are an estimated 2.5 million pressure ulcers/injuries per year in acute care hospitals in the United States, with a cost of $9.1 billion to $11.6 billion (Reddy, Gill, & Ronchon, 2006; Shreve, Van Den Bos, Gray, Halford, Rustagi, & Ziemkiewicz, 2010; Institute for Healthcare Improvement, 2014). In addition to increasing health care resource consumption and costs, pressure ulcers also cause pain to the patient, prolong hospital stays, and place patients at risk for other adverse events (Gorecki et al., 2009; Lyder et al., 2012; National Pressure Ulcer Advisory Panel & European Pressure Ulcer Advisory Panel, 2009). The occurrence of pressure ulcers is considered a serious consequence of substandard quality of care.

The prevention of pressure ulcers/injuries has become the focus of national policy and patient safety initiatives. NQF (2008) considers HAPUs of stages III and IV ‘‘largely preventable, grave errors’’ (p. 1). On October 1, 2008, CMS stopped reimbursing hospitals for costs of treating stage III and IV HAPUs (CMS, 2007; Stone et al., 2010). Additionally, CMS is planning to implement the Hospital-Acquired Condition (HAC) Reduction Program in the near future, under which hospitals will be penalized for excess rates of HAPUs and other HACs (CMS, 2014). National health care stakeholders, including the National Quality Strategy and the CMS Partnership for Patients and HAC Reduction Program, have identified pressure ulcers as a patient safety concern.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
Center for Medicare & Medicaid Services. (2007). FY 2008 inpatient prospective payment system final rule. Retrieved from http://www.cms.gov/Newsroom/MediaReleaseDatabase/Fact-sheets/2007-Fact-sheets-items/2007-08-012.html.

Center for Medicare & Medicaid Services. (2014). Fact sheets: CMS proposals to improve quality of care during hospital inpatient stays. Retrieved August 24, 2014, from http://www.cms.gov/Newsroom/MediaReleaseDatabase/Fact-sheets/2014-Fact-sheets-items/2014-04-30-2.html.

Gorecki, C., Brown, J. M., Nelson, E. A., Briggs, M., Schoonhoven, L., Dealey, C., … Nixon, J. (2009). Impact of pressure ulcers on quality of life in older patients: A systematic review. Journal of the American Geriatrics Society. 57(7), 1175–1183.

Institute for Healthcare Improvement. (2014). Protecting 5 million lives from harm: Overview. Cambridge, MA. Retrieved September 27, 2014, from http://www.ihi.org/engage/Initiatives/completed/5MillionLivesCampaign/Pages/default.aspx.

Lyder, C. H., Wang, Y., Metersky, M., Curry, M., Kliman, R., Verzier, N. R., & Hunt, D. R. (2012). Hospital-acquired pressure ulcers: Results from the national Medicare Patient Safety Monitoring System Study. Journal of the American Geriatrics Society, 60(9), 1603–1608.

National Pressure Ulcer Advisory Panel (NPUAP)/EPUAP. (2009). Prevention and treatment of pressure ulcers: Clinical practice guideline. Washington, DC: NPUAP.

National Quality Forum. (2008). Serious reportable events. Retrieved from https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=57355.

Reddy, M., Gill, S. S., & Rochon, P. (2006). Preventing pressure ulcers: A systematic review. The Journal of the American Medical Association, 296(8), 974–984.

Shreve, J., Van Den Bos, J., Gray, T., Halford, M., Rustagi, K., & Ziemkiewicz, E. (2010). The economic measurement of medical errors. Sponsored by Society of Actuaries' Health Section. Milliman Inc.

Stone, P. W., Glied, S. A., McNair, P. D., Matthes, N., Cohen, B., Landers, T. F., & Larson, E. L. (2010). CMS changes in reimbursement for HAIs. Medical Care, 48, 433–439. doi: 10.1097/MLR. 0b013e3181d5fb3f.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
Not applicable.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria 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. **Cross Cutting Areas** *(check all the areas that apply):*
Safety

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.)*
None at this time.

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

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** PAPUI_Data_Collection_Code_Sheet-635987554553524645.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
This is a new measure.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the*

target population, i.e., cases from the target population with the target process, condition, event, or outcome)
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
The total number of participants enrolled during the quarter that have at least one documented PU (of any stage) acquired while a PACE participant.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Quarterly data.

S.6. **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, 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.
Inclusion criteria for numerator:
• Include participants living at home or in assisted living facilities.
• Include participants with pressure injuries that developed and were identified less than 24 hours after the participant was in an emergency room, admitted to the hospital, nursing home, skilled nursing facility, hospice facility, or rehabilitation facility.

Exclusion criteria for numerator:
• Exclude participants who were not enrolled in a PACE Program for at least one day during the quarter.
• Exclude participants who were not in their home setting for at least one day of the quarter. For each participant, exclude participants who were only:
  o In a nursing home facility
  o In a hospice facility
  o In hospice care at home
  o In skilled nursing care, or
  o In a rehabilitation setting
• Exclude participants whose pressure ulcer/injury was acquired before they were enrolled in PACE.
• Exclude participants with other kinds of skin breakdown that developed during the quarter, such as diabetic ulcers or venous ulcers.
• Exclude participants whose only skin breakdown was documented as a "Kennedy Terminal Ulcer" during the quarter. Kennedy Terminal Ulcers are not acknowledged as a pressure ulcer/injury stage by NPUAP.
• Exclude participants with pressure ulcer/injury that developed and were identified less than 24 hours after a participant returned home (or to an assisted living facility).

Specific data collection items and responses:
• Participant No.
• Age (at end of month):
- Age in years if 55–89
- Age greater >89 = 90+
- Unknown = 99
• Gender:
- Male = 1
- Female = 2
- Unknown = 99
• Pressure Injury No.
• Month

- January = 1
- February = 2
- Etc.

• Pressure Injury Stage
- Stage I = 1
- Stage II = 2
- Stage III = 3
- Stage IV = 4
- Unstageable = 5
- Deep Tissue = 6
- Unknown = 99

Pressure Injury as defined by the National Pressure Ulcer Advisory Panel*:

A pressure injury is localized damage to the skin and/or underlying soft tissue usually over a bony prominence or related to a medical or other device. The injury can present as intact skin or an open ulcer and may be painful. The injury occurs as a result of intense and/or prolonged pressure or pressure in combination with shear. The tolerance of soft tissue for pressure and shear may also be affected by microclimate, nutrition, perfusion, co-morbidities and condition of the soft tissue.

Pressure ulcers/injuries are characterized by stage:

Stage 1 Pressure Injury: Non-blanchable erythema of intact skin
Intact skin with a localized area of non-blanchable erythema, which may appear differently in darkly pigmented skin. Presence of blanchable erythema or changes in sensation, temperature, or firmness may precede visual changes. Color changes do not include purple or maroon discoloration; these may indicate deep tissue pressure injury.

Stage 2 Pressure Injury: Partial-thickness skin loss with exposed dermis
Partial-thickness loss of skin with exposed dermis. The wound bed is viable, pink or red, moist, and may also present as an intact or ruptured serum-filled blister. Adipose (fat) is not visible and deeper tissues are not visible. Granulation tissue, slough and eschar are not present. These injuries commonly result from adverse microclimate and shear in the skin over the pelvis and shear in the heel. This stage should not be used to describe moisture associated skin damage (MASD) including incontinence associated dermatitis (IAD), intertriginous dermatitis (ITD), medical adhesive related skin injury (MARSI), or traumatic wounds (skin tears, burns, abrasions).

Stage 3 Pressure Injury: Full-thickness skin loss
Full-thickness loss of skin, in which adipose (fat) is visible in the injury and granulation tissue and epibole (rolled wound edges) are often present. Slough and/or eschar may be visible. The depth of tissue damage varies by anatomical location; areas of significant adiposity can develop deep wounds. Undermining and tunneling may occur. Fascia, muscle, tendon, ligament, cartilage and/or bone are not exposed. If slough or eschar obscures the extent of tissue loss this is an Unstageable Pressure Injury.

Stage 4 Pressure Injury: Full-thickness skin and tissue loss
Full-thickness skin and tissue loss with exposed or directly palpable fascia, muscle, tendon, ligament, cartilage or bone in the injury. Slough and/or eschar may be visible. Epibole (rolled edges), undermining and/or tunneling often occur. Depth varies by anatomical location. If slough or eschar obscures the extent of tissue loss this is an Unstageable Pressure Injury.

Unstageable Pressure Injury: Obscured full-thickness skin and tissue loss
Full-thickness skin and tissue loss in which the extent of tissue damage within the injury cannot be confirmed because it is obscured by slough or eschar. If slough or eschar is removed, a Stage 3 or Stage 4 pressure injury will

be revealed. Stable eschar (i.e. dry, adherent, intact without erythema or fluctuance) on an ischemic limb or the heel(s) should not be removed.

Deep Tissue Pressure Injury: Persistent non-blanchable deep red, maroon or purple discoloration
Intact or non-intact skin with localized area of persistent non-blanchable deep red, maroon, purple discoloration or epidermal separation revealing a dark wound bed or blood filled blister. Pain and temperature change often precede skin color changes. Discoloration may appear differently in darkly pigmented skin. This injury results from intense and/or prolonged pressure and shear forces at the bone-muscle interface. The wound may evolve rapidly to reveal the actual extent of tissue injury, or may resolve without tissue loss. If necrotic tissue, subcutaneous tissue, granulation tissue, fascia, muscle or other underlying structures are visible, this indicates a full thickness pressure injury (Unstageable, Stage 3 or Stage 4). Do not use DTPI to describe vascular, traumatic, neuropathic, or dermatologic conditions.

* This PU/I data collection will follow the NPUAP pressure ulcer/injury definition and staging categories. More information can be found in this link: http://www.npuap.org/national-pressure-ulcer-advisory-panel-npuap-announces-a-change-in-terminology-from-pressure-ulcer-to-pressure-injury-and-updates-the-stages-of-pressure-injury/

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Number of participants on a PACE organization's census during the quarter.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk : Dual eligible beneficiaries, Populations at Risk : Individuals with multiple chronic conditions, Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
Number of participants on the PACE site census at least one day during the quarter.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Exclude persons who were not on the PACE census for at least one day during the quarter. Exclude participants who lived outside their home/assisted living setting for every day of the quarter.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
• Exclude participants who were not on a PACE organization's census for each one day during the quarter.
• Exclude participants who were not in their home setting every day of the quarter. Exclude participants who spent the entire quarter living:
- In a nursing home facility
- In a hospice facility
- In hospice care at home
- In skilled nursing care, or
- In a rehabilitation setting

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
Risk stratification will be used rather than risk adjustment. Stratification will be based on PACE organization characteristics. Because PACE participants are frail elderly in each organization, they may be considered a single

population, not requiring risk adjustment to account for different populations across PACE organizations.

Two demographic variables—age and gender—will be collected so that the potential for sociodemographic adjustment can be assessed.
• Age is defined as the participant age at the end of the reporting month. It is to be recorded in single years from 55 through 89. To comply with HIPAA requirements, all participants aged 90 and above will be top coded at 90.
• Gender is to be classified as male or female.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Stratification by risk category/subgroup
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Not applicable.

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Provided in response box S.15a

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
Not applicable.

S.16. **Type of score:**
Ratio
If other:

S.17. **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.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
1. The target population is all included participants on a PACE organization's census for at least one day during a calendar quarter.
2. The numerator is the number of PACE participants whose clinical records documented the presence of one or more included pressure injuries during the quarter.
3. Count the number of included PACE participants on a PACE organization's census for at least one day during a calendar quarter.
4. Divide the quarterly number of participants with pressure injuries by the number of participants on the census during the quarter.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on*

minimum sample size.)
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
No sampling. Data to be collected from all PACE participants, subject to the exclusions listed above.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not applicable.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Pressure injury data are collected by month so that the impact of missing data can be reduced. PACE sites that fail to report data for 1 month, the same month for both the numerator and denominator, will have their quarterly rates based on 2 months of data. PACE programs that fail to report data for 2 months out of the quarter will not have rates calculated, as a 1-month sample decreases the reliability and potentially the validity of the data to an unacceptably low level.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Management Data, Paper Medical Records

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Collection instrument is provided as an uploaded appendix.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Other
If other: PACE programs provide services to participants who live in their own homes (or in home-like settings) in the community. Participants attend PACE centers regularly (e.g., 3 days per week) for a variety of activities and support services. If a participan

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
Not applicable.

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
PAPUI_testing_attachment_NQF-635987554004164141.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: PACE-Acquired Pressure Ulcer/Injury Prevalence Rate
Date of Submission: 5/13/2016
Type of Measure:

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

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed

performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are

different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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.23***)	**Measure Tested with Data From:**
☒ abstracted from paper record	☒ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☒ abstracted from electronic health record	☒ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

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

1.3. What are the dates of the data used in testing? January-February, 2015

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.26*)**	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☒ other: PACE Organization	☒ other: PACE Organization

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*)

For reliability, a sample of 50 sites was randomly selected out of a total of 114 PACE Organizations (POs). Additionally, the oldest and two newest PACE sites were included in the sample. A total of 29 of these sites submitted data from January and February 2015 for pressure ulcer/injury prevalence, with a minimum monthly census size of 1 participant to a maximum of 863 participants (mean = 198). One (1) of the sites had only one (1) participant in January and February 2015, and this participant had PACE-acquired pressure ulcers/injuries (PAPU/Is) but no stage 3 or above pressure ulcers/injuries. This site was excluded from the summary and analysis as it is an extreme outlier that cannot provide reliable pressure ulcer/injury rates. Characteristics of the PACE Organizations which submitted data are shown in Table 1 below.

Table 1. Characteristics of PACE Organizations Participating in Testing

Characteristic	*N (%)*
Affiliated With Academic Medical Center	
Yes	3 (10.7)
No	25 (89.3)
Location	
Metropolitan	22 (78.6)
Micropolitan	2 (7.1)
Non-metropolitan	4 (14.3)

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)

A random sample of PACE sites was asked to submit pressure ulcer/injury data for all of their participants. A total of 5,730 participants were included. Data on pressure ulcers/injuries was abstracted from health records.

Data on participant characteristics was not obtained.

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.

There were two test populations: (1) a set of experts for the content validity study and (2) a random sample of PACE sites for the reliability testing.

For content validity, a sample of 17 academic pressure ulcer/injury experts were identified, and 8 completed our validity testing survey. For reliability, a sample of 50 sites was randomly selected out of a total of 114 PACE Organizations (POs). Additionally, the oldest and two newest PACE sites were included in the sample. A total of 29 of these sites submitted data from January and February 2015 for pressure ulcer/injury prevalence. One (1) of the sites had only one (1) participant in January and February 2015, and this participant had PACE-acquired pressure ulcer/injury but no stage 3 or above pressure ulcers/injuries.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

No sociodemographic variables were collected or analyzed.

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*)

Site-level reliability of each measure was assessed using the signal-to-noise analysis approach. This approach was originally proposed by Adams for normally distributed data using a mixed

model (Adams, 2009). The assessment of reliability for the pressure ulcer/injury rate will use a similar approach modified for binary outcomes.

The signal-to-noise analysis, which is appropriate for the measures, quantifies the amount of variation in performance due to differences in sites (signal), as opposed to differences due to random variation within each site (noise). The signal-to-noise method results in a reliability statistic that ranges from 0 to 1 for each site. A value of 0 indicates that all variation is due to random variation, and a value of 1 indicates that all variation is due to real differences in site performance. The signal-to-noise approach for reliability assessment depends on the normality assumption for the distributions of these rates. For PACE-acquired pressure ulcer/injury rates, the distributions are not normal. One sites with fewer than 20 participants reviewed for pressure ulcers/injuries was excluded from analysis.

Citation:
Adams J. L. (2009). The reliability of provider profiling: a tutorial. RAND Corporation, Santa Monica.

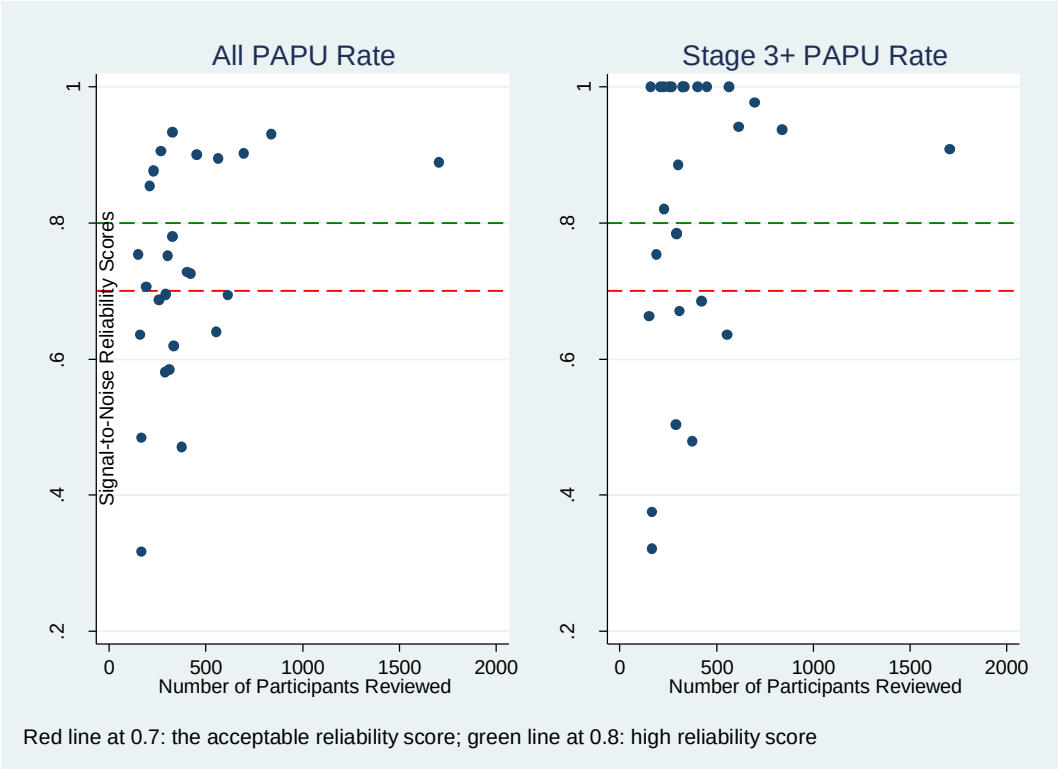
2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Table 2 summarizes the reliability assessment results. Regarding PACE-Acquired Pressure Ulcer/Injury (PAPU/I) rates, the mean reliability score for all PAPU/I rates was 0.73, while that for PAPU/I stage 3+ rates was 0.83. The distributions of the PAPU/I rates and PAPU/I stage 3+ rates are both skewed toward the right (Figure 1).

Table 2. Signal-to-Noise Assessment of Reliability of Pressure Ulcer/Injury Rates

	Reliability Scores	
Measures	**Mean (SD)**	**Median (Min, Max)**
PAPU/I rate (*n*=28)	0.73 (0.16)	0.73 (0.32, 0.93)
PAPU/I stage 3+ rate (*n*=28)	0.83 (0.21)	0.92 (0.33, 1.00)

Figure 1. Signal-to-Noise Reliability Assessment of PAPU/I Rates



Red line at 0.7: the acceptable reliability score; green line at 0.8: high reliability score

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 mean reliability score for all PAPU/I rates was acceptable at 0.73, while that for PAPU/I stage 3+ rates was high at 0.83. The distributions of the PAPU/I rates and PAPU/I stage 3+ rates are both skewed toward the right.

Two months of data were collected for the reliability study. PAPU/I rates measured over a longer period of time are needed to produce more reliable results for smaller PACE organizations. Among all sites, the median PAPU/I rate was 1.44, with a minimum of 0.31 and a maximum of 5.60, while the median for the PAPU/I stage 3+ rate was 0.38, with a minimum of 0 and a maximum of 4.57.

2b2. VALIDITY TESTING

2b2.1. 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*)

2b2.2. 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)*

Content validity was assessed using a national panel of pressure ulcer/injury experts to quantify experts' assessments of the validity of the PAPI numerator, denominator, and calculated rate. Content validity of the measure was analyzed by calculating item-level content validity indices (I-CVIs). The I-CVI indicates the proportion of experts who consider the item as content valid. Experts rated each component's content/face validity using a 4-point scale: 1 = very low (major modification needed), 2 = low (some modification needed), 3 = high (no modification needed but could be improved with minor changes), and 4 = very high (no modification needed). I-CVI is computed for each item by counting the number of experts giving a rating of 3 or 4 and dividing the number by the total number of experts (Polit, Beck, & Owen, 2007).

Citation:
Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health. 30*, 459-467.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Table 3: Content Validity Results for Data Elements in the PAPU/I Data Collection Instructions

Data Element	I-CVI
PAPU/I Prevalence Rate distinguishes good from poor quality of care	0.75 (6/8)
The measure captures what this measure intends to measure:	
PAPU/I Rate	0.88 (7/8)
PAPU/I Numerator	0.88 (7/8)
PAPU/I Denominator	0.88 (7/8)
Exclusions from both Numerator and Denominator:	
Each day of quarter participant not enrolled	0.88 (7/8)
Each day of quarter participant not in home setting	
• Hospitalized more than 23 hours	1.00 (8/8)
• In emergency room more than 23 hours	0.88 (7/8)
• In a nursing home facility	1.00 (7/7)
• In a hospice facility	0.88 (7/8)
• In hospice care at home	1.00 (8/8)
• In skilled nursing care	1.00 (8/8)
• In a rehabilitation setting	1.00 (8/8)
Exclusion Criteria for Numerator:	
Pressure ulcer/injury acquired before PACE enrollment	1.00 (8/8)
Other kinds of skin breakdown that developed during the quarter (e.g. diabetic ulcers, venous ulcers)	0.75 (6/8)
Kennedy Terminal Ulcers	0.63 (5/8)

Pressure ulcers/injuries that developed and were identified less than 24 hours after a participant returned home (or to an assisted living)	0.86 (6/7)
Exclusion Criteria for Denominator:	
Deceased participants after the date of death	1.00 (7/7)
Inclusion Criteria for both Numerator and Denominator:	
Participants in assisted living facilities	1.00 (8/8)
Inclusion Criteria for Numerator:	
Pressure ulcers/injuries that developed and identified less than 24 hours after the participant was in emergency room, admitted to the hospital, nursing home, skilled nursing facility, hospice facility, or rehabilitation facility	0.83 (5/6)
Inclusion Criteria for Denominator:	
Each day a participant was on the participant census after enrolling	1.00 (7/7)

2b2.4. 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?*)

Polit et al. (2007) suggested that items with good content (or face) validity should have an I-CVI of .78 or higher from three or more experts' review. Based on this, we used .78 as a cutoff point to determine good, acceptable content (or face) validity. Another evaluation criterion was based on Lynn (1986). Lynn (1986) argued that the disagreement is accepted only if "six or more experts" gave an item a rating of 1 (very low) or 2 (low).

A total of 8 academic experts completed content validity testing. As shown in Table 2 above, the majority of items on the content validity testing survey had good validity as indicated by an I-CVI of greater than 0.78 (16 of 20 items or 75%). In addition, none of the items was disagreed upon by 6 or more experts.

Citations:

Lynn, M. (1986). Determination and quantification of content validity. *Nursing Research, 35,* 381–385.

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Research in Nursing & Health. 30, 459-467.

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

2b3.1. 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*)

Content validity of the measure exclusions was analyzed by calculating item-level content validity indices (I-CVIs). The I-CVI indicates the proportion of experts who consider the item as content valid. Experts rated each component's content/face validity using a 4-point scale: 1 = very low (major modification needed), 2 = low (some modification needed), 3 = high (no

modification needed but could be improved with minor changes), and 4 = very high (no modification needed). I-CVI is computed for each item by counting the number of experts giving a rating of 3 or 4 and dividing the number by the total number of experts (Polit, Beck, & Owen, 2007).

Polit et al. (2007) suggested that items with good content (or face) validity should have an I-CVI of .78 or higher from three or more experts' review. Based on this, we used .78 as a cutoff point to determine good, acceptable content (or face) validity. Another evaluation criterion was based on Lynn (1986). Lynn (1986) argued that the disagreement is accepted only if "six or more experts" gave an item a rating of 1 (very low) or 2 (low).

Citations:

Lynn, M. (1986). Determination and quantification of content validity. *Nursing Research, 35,* 381–385.

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Research in Nursing & Health. 30, 459-467.

2b3.2. 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*)

A total of 8 academic experts completed content validity testing. As shown in Table 2 above, three of the five exclusion criteria (pressure ulcers/injuries acquired before enrollment in PACE, pressure ulcers/injuries acquired less than 24 hours after return to home, and for the denominator, excluding participants who died during the reporting period) had good content validity with I-CVIs of 1.00, 0.86, and 1.00 respectively. The other two exclusion criteria (other types of skin breakdown and Kennedy Terminal Ulcers) did not have acceptable validity scores.

2b3.3. 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 exclusions are needed to produce valid pressure injury rates. Furthermore, we agree with the experts who thought that other types of skin breakdown and Kennedy Terminal Ulcers should be excluded, even though the validity analysis did not find those two criteria to have good validity.

2b4. 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 2b5.

2b4.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**

☐ **Statistical risk model with** Click here to enter number of factors **risk factors**
☒ **Stratification by 2 risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use 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.

We do not see a need for risk adjusting at the participant level, as there is a presumptive reason to believe that PACE participants have a high degree of homogeneity as frail elderly who have been determined to be nursing home-eligible.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

Risk stratification will be used rather than risk adjustment. Stratification will be based on PACE site characteristics. Because PACE participants are frail elderly in each site, they may be considered a single population, not requiring risk adjustment to account for different populations across PACE sites.

After implementation two demographic variables—age and gender—will be collected so that the potential for sociodemographic adjustment can be assessed.

- Age is defined as the participant age at the end of the reporting month. It is to be recorded in single years from 55 through 89. To comply with HIPAA requirements, all participants aged 90 and above will be top coded at 90.
- Gender is to be classified as male or female.

We conducted correlations to determine the associations between PACE site PAPU/I rates with mean site-level age of those with PAPU/I and site-level proportion of males with PAPU/I, respectively. Additionally, we calculated correlation coefficients for rates of PACE site Stage 3+ PAPU/I rates with mean age site-level age of those with Stage 3+ PAPU/I and site-level proportion of males with Stage 3+ PAPU/I.

2b4.4a. What were the statistical results of the analyses used to select risk factors?

Table 4: Characteristics of Participants With Pressure Ulcers and Characteristics of Pressure Ulcers/Injuries

Pressure Ulcers/Injuries		
Participant-Level Characteristics (n = 219)	Summary Statistics	
Age	Mean (SD)	77.41 (9.63)

Gender	Male Female	60 (27.40%) 159 (72.60%)
Number of pressure ulcers/injuries	Mean (SD) 0 1 2+	1.51 (1.40) 56 (25.60%) 116 (53.00%) 47 (21.50%)
Stage of pressure ulcers/injuries	Stage I Stage II Stage III Stage IV Unstageable, sDTI Unstageable, Other Unknown	17 (7.33%) 118 (50.86%) 30 (12.93%) 15 (6.47%) 30 (12.93%) 19 (8.19%) 3 (1.29%)

Figure 2: Correlation Between All Stages of PAPU/I Rates and Mean Age (*r* = –0.16, *n* = 24)

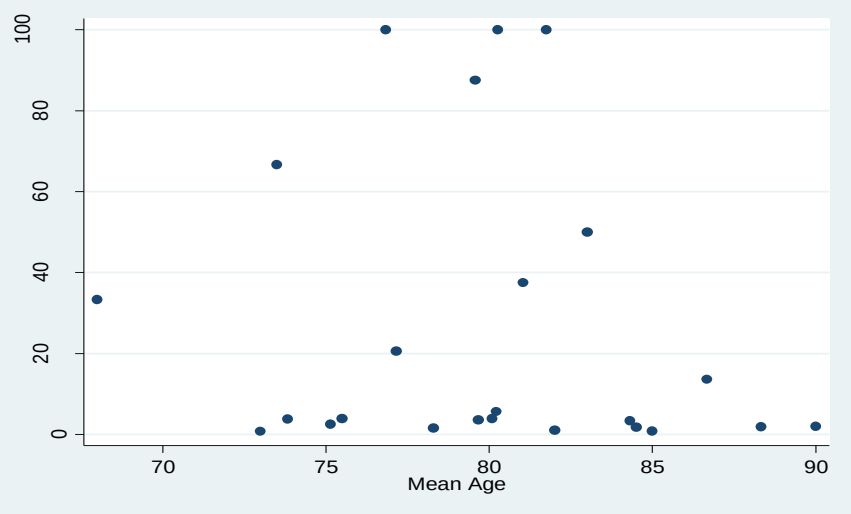


Figure 3: Correlation Between Stage 3+ PAPU/I Rates and Mean Age (*r* = –0.20, *n* = 20).

Mean age was calculated by site-level mean age of participants having Stage 3+ PAPU/Is.

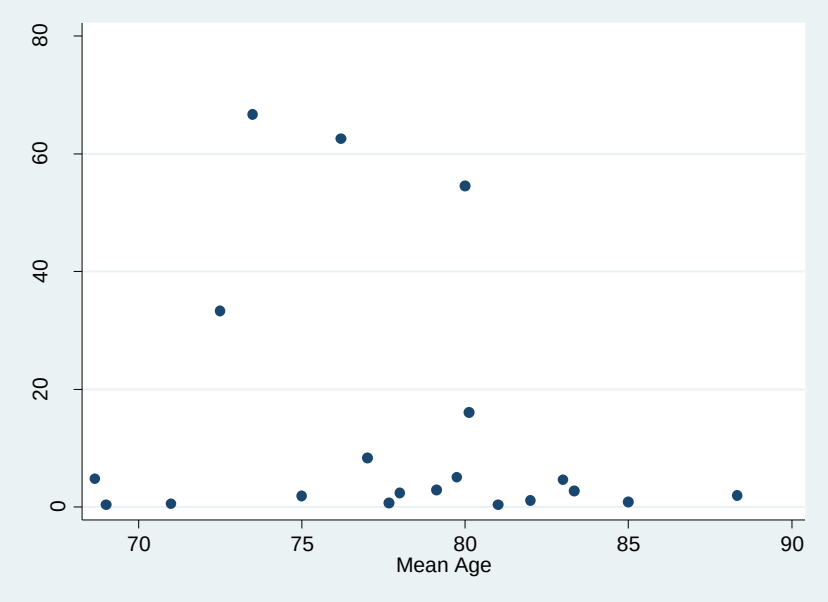
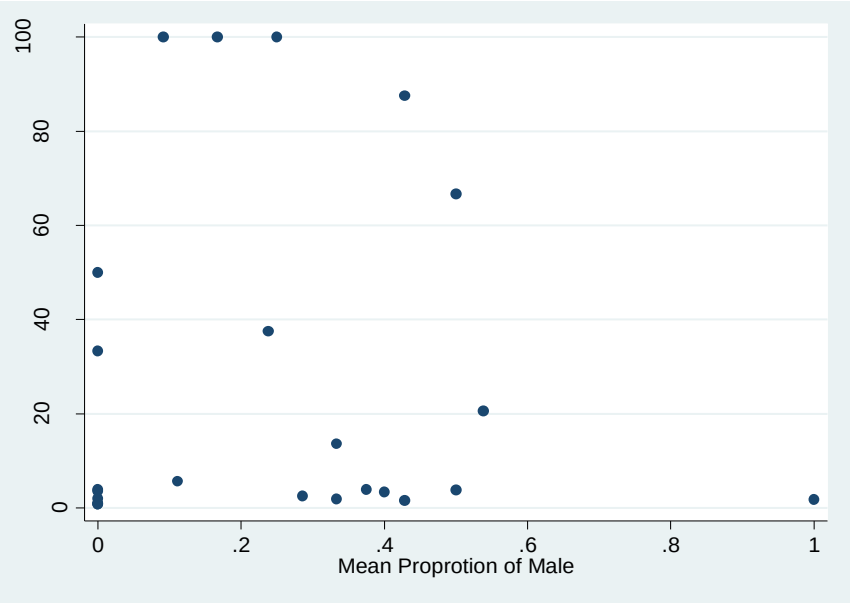


Figure 4: Correlation Between All PAPU Rates and Mean Proportion of Male (*r* = –0.03, *n* = 24).

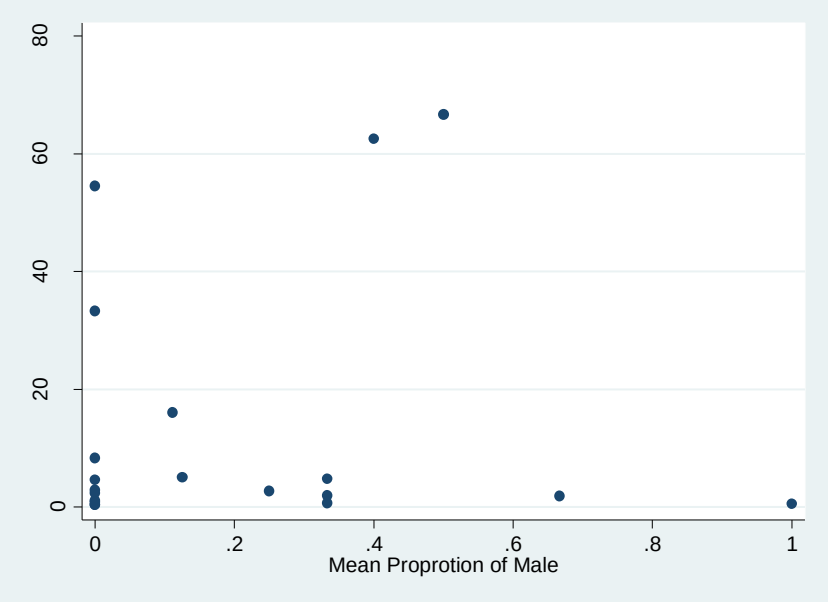
Mean proportion of male was calculated by site-level mean proportion of male having PAPUs. Negative correlation indicates that sites having more males with PAPUs were likely to have lower rates of all PAPU rates.



Note: One PACE site was excluded from scatterplot because of outlier data.

Figure 5: Correlation Between Stage 3+ PAPU Rates and Mean Proportion of Male (r = 0.09, n = 20).

Mean proportion of male was calculated by site-level mean proportion of male having Stage 3+ PAPUs. Positive correlation indicates that sites having more males with Stage 3+ PAPUs were likely to have lower rates of Stage 3+ PAPU rates.



Note: One PACE site excluded from scatterplot because of outlier data.

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

All-stage PAPU/I rates and Stage 3+ PAPU/I rates both had weak negative correlations with mean age (r = –0.16 and r = –0.20). Further, both all-stage PAPU/I and Stage 3+ PAPU/I rates were very weakly/negligibly correlated with gender (r = -0.03 and r = 0.09).

After implementation of the measure, we will continue to collect data to determine the usefulness of risk-stratification based on age and gender.

2b4.5. 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 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

2b4.10. 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*)

2b4.11. 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*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. 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)*

Due to our small sample size, we did not conduct statistical analyses to determine differences in performance across PACE sites. However, the descriptive statistics indicate that there are differences in PAPU/I rates per 100 participants across PACE sites (mean = 3.67, SD = 32.25, median = 3.67, range = 0-100). After implementation, we will conduct further analyses to determine significant differences in PAPU/I rates across PACE sites.

2b5.2. 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*)

Not Applicable.

2b5.3. 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?*)

Not Applicable.

2b6. 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 SDS 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 SDS 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.***

Not Applicable.

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Not Applicable.

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

Not Applicable.

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

Not Applicable.

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

We had no missing data on the numerator or denominator for pressure ulcers from sites included in the sample.

2b7.2. 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)

Not applicable.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

Not applicable.

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), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

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)*

Some data elements are in defined fields in 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.**

Some PACE Organizations do not use electronic medical records. All organizations will abstract data manually for this measure from either their electronic or paper charts.

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.**

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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

We conducted a survey with PACE organizations to collect information on their experiences with data collection. Overall, the data collection time was reasonable, around 4 hours with less than an hour for data submission. While the sites reported a fairly high data collection burden, this was balanced by the fact that 60% of the sites stated that the data were very easy to obtain. Almost all (91%) of the sites stated that PAPU/I rates are useful for quality

improvement and over half felt that the rates would be a valid way to distinguish good from poor care quality. Thus, although there is a perceived data collection burden, this is outweighed by the usefulness of the data for quality improvement and distinguishing PACE sites based on their quality of care. Because of the high reported ease of obtaining the data, we anticipate that the perceived data collection burden will decrease as sites become more familiar with the data collection and submission process.

· 68% of PACE organizations participating in the study manually extracted pressure injury data from electronic health records. Just 2 of 22 PACE organizations collected pressure ulcer data from paper records.
· The median time required for pressure injury data collection was approximately four (4) hours. Data submission took less than one hour. 57% of PACE organizations categorized the data collection burden as high or very high.
· 60% of PACE organizations said that it was "very easy" to obtain the data elements for pressure injuries.
· 91% of PACE organizations in the study "agreed" or "strongly agreed" that pressure injury data was useful for quality improvement. 57% said that pressure injury rates would be a valid way to distinguish good from poor quality of care at PACE sites.
· Reporting for auxillary data (risk assessment, prevention activities) related to pressure injuries was low. Most PACE organizations could not find such data elements in their systems

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
None.

4. Usability and Use

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

4a. Accountability and Transparency

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

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

Planned	Current Use (for current use provide URL)
Public Reporting Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Not in use	

4a.1. **For each CURRENT use, checked above, provide:**

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

Not applicable.

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
This is a new measure. CMS is evaluating its use in upcoming PACE quality programs.

4a.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 is considering the use of the PACE-Acquired Pressure Ulcer/Injury Prevalence Rate in accountability applications within the next two years.

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

Not applicable as this is a newly developed measure.

4b.2. **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.**
Not applicable as this is a newly developed measure.

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
No negative unintended consequences have been identified.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure

focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures

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

Yes

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

0201 : Pressure ulcer prevalence (hospital acquired)
0538 : Pressure Ulcer Prevention and Care
0678 : Percent of Residents or Patients with Pressure Ulcers That Are New or Worsened (Short-Stay)
0679 : Percent of High Risk Residents with Pressure Ulcers (Long Stay)

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

Not applicable.

5a. Harmonization

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 completely harmonized?

No

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

The measures being developed for the PACE program are not closely aligned with any of the four endorsed pressure ulcer/injury measures. It appears that they all use the same conceptual definition of a pressure ulcer/injury, although the data sources and methods differ enough from each other to result in concrete definitional differences. In addition to differences in data sources, none of the related measures collect data on pressure injuries acquired in the home setting or pressure ulcers/injuries in PACE participants. The proposed measure includes pressure injuries of any stage in PACE participants. Percent of High-Risk Residents With Pressure Ulcers (Long Stay) (NQF 0679) is limited to high risk long-stay patients in nursing facilities with pressure ulcers that are Stage II or greater, while Percent of Residents or Patients With Pressure Ulcers That Are New or Worsened (Short Stay) (NQF 0678) is limited to short-stay nursing facility patients with Stage II–IV pressure ulcers that are new or worsened since the prior assessment. Pressure Ulcer Prevalence (Hospital Acquired) (NQF 0201) is limited to pressure ulcers Stage II or greater acquired during a stay in an acute care hospital, and Pressure Ulcer Rate (NQF 0538) is limited to pediatric hospitals.

5b. Competing Measures

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

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

Not applicable.

Appendix

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

Attachment **Attachment:** AppendixA1_PAPUI_Data_Collection_Sheet.docx

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** CMS
Co.2 **Point of Contact:** Stacy, Davis, stacy.davis@cms.hhs.gov, 410-786-7813-
Co.3 **Measure Developer if different from Measure Steward:** Econometrica, Inc.
Co.4 **Point of Contact:** Mark, Stewart, mstewart@econometricainc.com, 240-204-5168-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
The PACE-Acquired Pressure Ulcer/Injury Prevalence Rate measure was developed in partnership with CMS by a team lead by Econometrica, Inc. consisting of Econometrica (prime contractor); the University Of Kansas Medical Center Research Institute (KUMCRI; subcontractor); Drs. Rosemary Kennedy and Barbara Resnick, and Ms. Heidi Bossley (consultants).

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:**
Ad.3 **Month and Year of most recent revision:**
Ad.4 **What is your frequency for review/update of this measure?**
Ad.5 **When is the next scheduled review/update for this measure?**

Ad.6 **Copyright statement:**
Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

A.1 Appendix PACE-Acquired Pressure Ulcer/Injury Prevalence Rate

PACE-Acquired Pressure Ulcer/Injury Data Abstraction Sheet

1/31/2015

Total Number of Participants on Census on the last day of the month	

If there were no participants with PACE-acquired pressure ulcers/injuries at your site in a month, enter 0 (zero) on the

Participant with PAPU/I	PAPU/I No.	Age (at end of month)	Gender	M
		Age in years if 55–89 90+=Age greater >89 99=Unknown	1=Male 2=Female 99=Unknown	1=Ja 2=F
001	1			
001	2			
002	1			
002	2			
003	1			
003	2			

Participant Census Days

January 2015	Number of Participants in Census
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	

NATIONAL QUALITY FORUM

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 3001
Measure Title: PACE Participant Fall Rate
Measure Steward: CMS
Brief Description of Measure: The quarterly incidence rate of falls amongst PACE participants per 1,000 participant days.
Developer Rationale: Fall Rates have been found to be an important safety concern in acute care and long-term care settings. There is evidence that falls are one of the most common adverse patient events in hospitals, and they are a source of significant injury, disability, and/or death. Nearly one-third of community-dwelling individuals over age 65 fall each year (Currie, 2008). In 2013, this accounted for nearly 2.5 million injury falls—with nearly two-thirds of this number experienced by females (CDC, 2013). Several national health care organizations—including the National Quality Strategy, the Partnership for Patients, and the CMS Hospital-Acquired Condition (HAC) Reduction Program—have identified patient falls as a patient safety concern.

Every fall carries a risk of injury. Clinicians can reduce injuries in part by reducing the risk of falling. Focusing prevention efforts solely on falls with injury is a faulty approach for improving patient safety. To some extent, falls with injury are a function of patient frailty; by contrast, the total fall rate is not influenced by differences among patients' susceptibility to injury.

Many if not most falls may result in no injury or only minor injury. Nevertheless, any fall may result in emotional distress and increased risk of falling in the future. Preventing falls among the frail elderly contributes to the maintenance of the participant's functional status and place in the community and the prevention of costs of treatment associated with falls. It is important to monitor all falls, not just falls with injury.

Citations:
CDC. (2013, December). WISQARS. Retrieved December 1, 2014 from Leading Causes of Nonfatal Injury Reports, 2001–2013: http://www.cdc.gov/injury/wisqars/nonfatal.html.

Currie, L. (2008). Fall and Injury Prevention. In R. Hughes (Ed.). Patient Safety and Quality: An Evidence-Based Handbook for Nurses. Rockville, MD: AHRQ. Retrieved November 18, 2014 from http://www.ncbi.nlm.nih.gov/books/NBK2653/.

Numerator Statement: Falls experienced by Participants in the PACE program during the month.
Denominator Statement: The denominator represents exposure of PACE participants to the risk of falling.

Denominator Exclusions: Exclude persons who were not enrolled as PACE participants, or who were not in their home location.
Measure Type: Outcome **Data Source:** Electronic Clinical Data : Electronic Health Record, Management Data, Paper Medical Records **Level of Analysis:** Facility

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

Summary of evidence:

- The developer provides the structural and process factors that influence fall rates and cites a few studies that find an indirect relationship between inpatient staffing and fall rates. The developer also calls out two studies that found, through a systematic review and meta-analysis, that fall prevention activities can reduce falls by up to 30 percent.

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*

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

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

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

- The developers collected data from a sample of 50 sites which were randomly selected out of a total of 114 PACE sites. A total of 34 of these sites submitted data from January –March 2015 for the fall rate. One site was excluded. They found a mean fall rate of 4.27 per 1,000 participant day (n=33). The mean rate appears to be higher that the rates obtained from primarily hospital-based studies provided by the developer after a review of the literature.

Disparities

- The developers examined fall rates based on two demographic variables, age and gender, to that the potential so socio-demographic adjustment could be assessed. Both PACE-site mean participant age and mean proportion of males had very weak correlations with total fall rates (r = 0.08 and r = -0.14, respectively).
- Several studies have demonstrated a difference in falls rates for specific populations. Disparities have been identified according to age, gender , disability, and race/ethnicity. Hospitalization for hip fractures due to falls is significantly higher for females than for males. However, fatality rates due to falls are higher for men than for women, and higher for Caucasians compared to African-

Americans. Among community-dwelling older women, age-adjusted fall rates are not different between African-Americans and Caucasians.

Questions for the Committee:

- *Specific question on information provided for gap in care.*
- *Is there a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*

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

Comments: **The developer presents a structure and process model that is supported by evidence that is predominantly hospital-based. This message is specific to participants in the PACE program who are living in the community - in their homes or assisted living. A body of literature does exist on structure and process variables that are more appropriate for the community based setting. For example: Hanley, A., Silke, C., & J. Murphy (2011). Community-based health efforts for the prevention of falls in the elderly. Clinical Interventions in Aging, 6, 19-25 (cited the importance of Occupational Therapy for environmental assessment and risk reduction); Pynoos, J., Steinman, B.A., & Nguyen, A. (2010). Environmental assessment and modification as fall-prevention strategies for the older adult. Clinics in Geriatric Medicine, 633-644; Wenjun, Li, et al., (2006) Outdoor falls among middle aged and older adults: a neglected public health problem. Am. Journal of Public Health, July, 96 (7) 1192-1200; Cochrane Reviews: Gillespie et al 2012;
**This is an outcome measure. Since there is no comparable setting to PACE, the developer could only rely upon evidence from settings such as hospitals and SNFs. PACE participants are served (as insurer and provider) in various settings, but mostly in their own homes in the community, making it most closely aligned with home care. The developer focused on two studies that found that fall prevention activities can reduce falls by up to 30 percent.
**Yes. literature review
**Yes- definitely pass

1b. Performance Gap

Comments: **The variability of structures for the PACE program are not included in the model for structure and process variables. The denominate for this fall rate is "exposure to the risk of falling", which is different that number of days in the PACE organization. These falls are among frail adults 55 yoa and older, who are living at home or in assisted living. Many structure and process variables are very different than in the hospital. In 2008, Mathematica Policy Research, Inc., reported an analysis of the effect of the PACE program on quality when compared with Medicaid home and community-based services (HBCS). PACE operated in 14 states then. PACE program like HBCS did not appear to prevent falls; but the PACE program had higher levels of preventive care. These variables should inform the structure and process variables for this measure's population and setting of care. And since PACE uses the IADL functional tool, this study found statistical significance that the PACE program was less likely to improve in getting around, which correlates with falls inside and outside the home. In this report, if a fall occurred the prior 6 months, the person's care was considered "unsuccessfully managed" (p. 42).
**This population is extremely vulnerable and the model presented for the gap in care is under-developed.
**Yes. And even if performance gap were narrow, there would be an argument to track falls because of their important contribution and documentation of declining strength and overall health in a vulnerable population.

**Yes it does. I would rate this high

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s):

Numerator:

A PACE participant fall is a sudden, unanticipated descent in which a participant comes to rest on the floor or some other surface, person, or object.

Inclusion Criteria:

- All PACE participant falls occurring in the participants home; in assisted living facilities, if that is their usual place of residence; in the PACE center, or in the care of a PACE transportation operator.
- Participants who are assisted to the floor by a care provider (assisted fall) are to be included in the count of falls.

Exclusion Criteria:

- Participants who fall (or sink) back to a bed, chair, car seat, walker seat, or toilet are excluded in the count of falls.
- Exclude falls in the participant home by staff, visitors, family members, or others who were not PACE participants
- Exclude participants who were not in their home location. For example, exclude participants who were in an emergency room, hospitalized, in a long term care facility, in a hospice facility, in skilled nursing care, in a rehabilitation setting.

Specific data collection items and responses:

- Fall Auto No.
- Month of Fall
  - January = 1
  - February = 2
  - Etc.
- Age (at end of month):
  - Age in years if 55–89
  - Age greater >89 = 90+
  - Unknown = 99
- Gender:
  - Male = 1
  - Female = 2
  - Unknown = 99

Denominator:

The denominator represents exposure of PACE participants to the risk of falling.

Questions for the Committee :

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

N/A

A sample of 50 sites was randomly selected out of a total of 114 PACE sites. Additionally, the oldest and two newest PACE sites were included in the sample. A total of 34 of these sites submitted data from January - March 2015 for the Fall Rate, with these sites having a minimum census size of 1, and a maximum of 854 participants (median = 190). There was one (1) large outlier with 24.79 falls per 1,000 participant days. This site has an unusually low number of participant days (121) because it had only one (1) participant in January and February 2015 and two (2) participants in March 2015. This site was excluded from the analysis and site-level descriptive statistics for total participant days and total fall rates.

Data were collected on a total of 1,995 falls. Many PACE organizations operate multiple program sites. To minimize respondent burden, organizations with multiple sites were instructed to report falls data for participants at their oldest site.

SUMMARY OF TESTING

Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☐ **Yes** ☐ **No**

Method(s) of reliability testing Signal-to-noise analysis

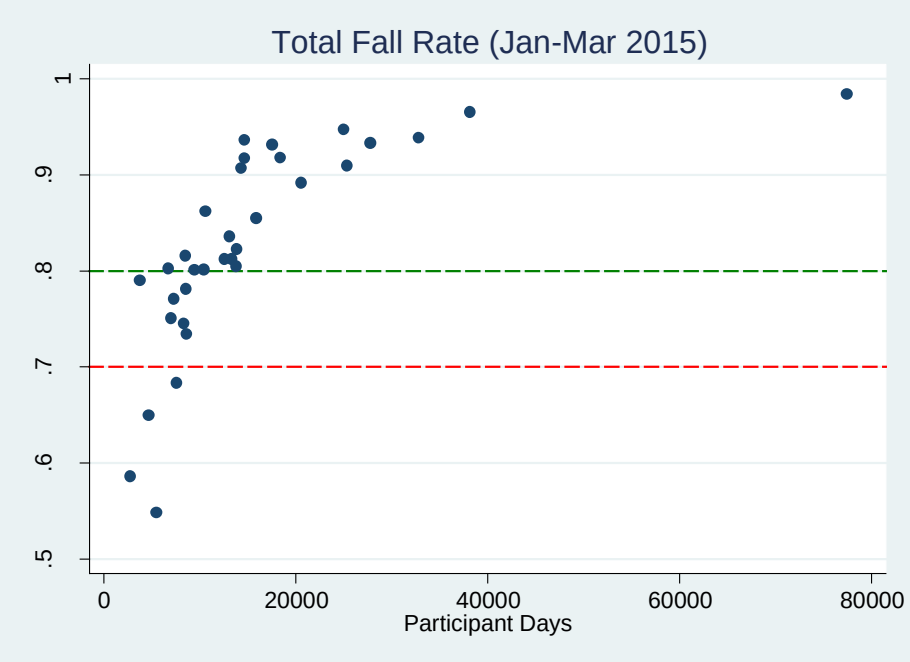
Results of reliability testing

Table 3 summarizes the reliability assessment results. The mean reliability score for the Fall Rate was 0.83. When we plotted the reliability scores versus the total participant days (Figure 1), there was a highly significant direct association between the reliability score of the Total Fall Rate and total participant days (r=0.66, p<0.001).

Table 1. Signal-to-Noise Assessment of Reliability of the 3-Month Total Fall Rate

Measures	Reliability Score: Mean (SD)	Median	Minimum, Maximum
Total participant fall rate (n=33)	0.83 (0.10)	0.82	0.55, 0.98

Figure 1. Signal-to-Noise Reliability Scores of Total Fall Rates of 33 PACE Sites



Note: A reliability score of 0.8 or higher is considered high, and a reliability score between 0.7 and 0.8 is considered acceptable.

Guidance from the Reliability Algorithm Based on the data presented by the developer, with a signal-to-noise ratio analysis, which averaged 0.83 but varied from 0.55-0.98, the reliability is judged as "MODERATE".

Questions for the Committee:

- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

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

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

○ *Are the specifications consistent with the evidence?*

2b2. Validity testing

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, summarize the validity testing from the prior review:

N/A

SUMMARY OF TESTING

Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

- ☒ **Face validity only**
- ☐ **Empirical validity testing of the measure score**

Validity testing method:

Content validity was assessed using a panel of 12 experts to: (1) quantify experts' degree of agreement regarding the content of the measure instructions (i.e., PACE Measure Instructions) and (2) obtain experts' narrative comments on the measure instructions. Based on the findings from content validity testing with expert review, we revised the measure. The revised instructions were distributed to PACE sample sites for pilot data collection.

Content and face validity of the measure was analyzed by calculating item-level content validity indices (I-CVIs). The I-CVI indicates the proportion of experts who consider the item as content (or face) valid. Experts rated content/face validity using a 4-point scale: 1 = very low (major modification needed), 2 = low (some modification needed), 3 = high (no modification needed but could be improved with minor changes), and 4 = very high (no modification needed). I-CVI is computed for each item by counting the number of experts giving a rating of 3 or 4 and dividing the number by the total number of experts (Polit, Beck, & Owen, 2007).

Validity testing results:

Table 4 displays I-CVIs for the Fall Rate measure.

Table 4:Content Validity Results From Experts for Data Elements in the PACE Participant Fall Rate

Fall Rate Data Element	I-CVI
Measure description	1.0 (6/6)
Definitions:	
• Numerator	1.0 (10/10)
• Denominator	.90 (9/10)
Measure calculation	.86 (6/7)
Exclusion criteria:	

- Falls by staff, visitors, or others who were not PACE participants.1.0 (2/2)

Overall applicability of the indicator to the PACE participants and PACE sites .92 (11/12)

Note: I-CVI = item-level content validity index; I-CVI/ave = average of I-CVIs. Each parenthesis indicates the number of experts who rated the data element as 3 or 4 divided by the total number of experts who responded.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality*

2b3-2b7. Threats to Validity

2b3. Exclusions:

This was not tested. According to the developer, "Excluding falls for people other than PACE participants and falls back to a bed or chair are relatively straightforward and don't warrant testing."

Questions for the Committee:

- *Is the developer's justification of no exclusions reasonable?*

2b4. Risk adjustment: **Risk-adjustment method** ☐ **None** ☐ **Statistical model** ☒ **Stratification**

The developer proposes risk stratification by 2 risk categories: Age and Gender. Stratification will be done based on PACE site characteristics.

Conceptual rationale for SDS factors included ? ☐ **Yes** ☒ **No**

SDS factors included in risk model? ☐ **Yes** ☒ **No**

Risk adjustment summary

The developer wrote, "Both mean age and mean percent male were weakly correlated with fall rates (r = 0.08 and r = -0.14, respectively). After implementation of the measure, we will continue to collect data to determine the usefulness of risk-stratification based on age and gender."

Questions for the Committee:

- *Is there sufficient evidence to suggest risk-stratification by age and gender, especially given such low correlations?*
- *Do you agree with the developer's decision, based on their analysis, to not include SDS factors in their risk-adjustment model?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

The developer wrote," Due to our small sample size, we did not conduct statistical analyses to determine differences in performance across PACE sites. However, the descriptive statistics indicate that there are differences in total fall rates across PACE sites (mean = 6.25, SD = 6.83, median = 3.93, range = 1.0-32.06). After implementation, we will conduct further analyses to determine significant differences in performance for total fall rates across PACE sites."

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

N/A

2b7. Missing Data

The developer wrote, "Because of the small amount of missing data, we did not conduct analyses of responders vs. nonresponders."

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

Committee pre-evaluation comments

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

2a1. & 2b1. Specifications

Comments: **The denominator is "exposure to the risk of falling" which is not assigned to denominators for other fall rates calculated in the articles reviewed by this developer. It would be expected that all PACE participants have exposure to fall risks in the home (inside / outside). The inclusion criteria is also expanded beyond the home / assisted living to also include falls that occur in the PACE center, or in the care of a transportation operator. These are not articulated in the structural model, and the context of each setting is very different.
**Face Validity was conducted by 12 members of an expert panel.
**The specifications are clearly defined. Exclusions include PACE participants who reside in a nursing home, hospital, or who are in an ED because falls are collected in these facilities, and there is a concern about double-counting. Specifications are consistent with the evidence provided.
**Reliability is clear. Definitions are clear and make sense.
**Yes the specifications are consistent. They are also meaningful to the PACE population.

2a2. Reliability Testing

Comments: **Signal to noise testing was conducted (variations due to performance differences in sites [signal] rather than differences due to random variation within each site (noise). However, the characteristics of each 34 sites is not described in terms of interdisciplinary teams, services provided etc. The reliability score was between 0.7 and 0.8, considered acceptable.
**Measure score reliability testing was done using signal-to-noise analysis. The mean score ws 0.83, with a range of 0.55 to 0.98.. There was a significant direct association between the reliability score of the total fall rate and total participant days.
**Looks OK
**Face validity only

2b2. Validity Testing

Comments: **Content validity (I-CVI) testing using a 12-member panel of experts
Measure description
Numerator Definition
Denominator Definition
Measure calculation
Overall acceptability 0.92
**Measure score testing was done using face validity only. A TEP of 12 people was used to quantify degree of agreement with the content of the measure instructions and to obtain comments.
**The measure is an indicator of quality based on the experience in settings such as hospitals, SNFs, IRFs, LTCHs and home care.
**Statistical model used here. Looks OK.
**There is no SDS risk adjustment in the measurement- uncertain of the co-morbidity risk assessment adjustment

either.

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **The fall rate excludes those participants who are not in the home environment - however, the fall can occur in the PACE center or awhile in the care of a transported. Also, falls back to a bed or chair are excluded as they are relatively straightforward and don's warrant testing. All falls should be included.
**Exclusions of falls for people (visitors, family, staff) other than PACE participants makes sense.
Risk stratification was done by age and gender. The developer intends to continue to collect data to determine the usefulness of risk-stratification based on age and gender.
Only descriptive statistics indicate differences in total fall rates across PACE sites. Small sample size did not allow for analysis to determine differences in performance across sites.
**Only enrolled members in PACE program are counted. I agree with the decision not to do risk adjustment based on SDS factors. The risk is that the risk adjustment would blur the results based on actual care factors.

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.

- This measure is generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score) Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)
- Some data elements are in defined fields in a combination of electronic sources.
- Some PACE Organizations do not use electronic medical records. All organizations will abstract data manually for this measure from either their electronic or paper charts.
- After collecting data from PACE sites for feasibility and reliability testing, a post-data collection survey was conducted, to ask PACE sites about data that they did not have available, data collection burden, and other issues.
- Some sites reported a fairly high data collection burden, however, this was balanced by the fact that over half of the sites stated that the data were very easy to obtain. Although there is a perceived data collection burden, this is outweighed by the usefulness of the data and comparative benchmarks.
- Because of the high reported ease of obtaining the data, we anticipate that the perceived data collection burden will decrease as sites become more familiar with the data collection and submission process.
- No fees or licensing requirements to use any aspect of the measure as specified, were reported.

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

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **Same response as for the PACE Fall Injury Rate
**The measure is generated by health care personnel in the rendering of care. Some data is available electronically, but many PACE organizations do not have EMRs. All PACE organizations will abstract data manually from either electronic or paper records. A high data collection burden was reported by the sites, but this was offset by the POs stating that the data were easy to obtain. This burden should decrease as familiarity with the data collection and submission process increases.
**Seems feasible. Only real problem would be failing to report a fall
**Data can be collected-

Criterion 4: Usability and Use

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

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

Current uses of the measure:

Publicly reported? ☐ **Yes** ☒ **No**

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

OR

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

Accountability program details:

- This is a new measure and is not currently in use.
- CMS is considering the use of the PACE Participant Fall Rate in accountability applications within the next two years.

Improvement results:

- Improvement data will be obtained once the measure has been implemented and tracked over time.

Unexpected findings (positive or negative) during implementation:

- No unexpected findings reported.

Potential harms:

- No negative unintended consequences have been identified.

Feedback :

- Developer did not identify any specific feedback loops related to this 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

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **To examine "all falls" as a quality improvement indicator lacks precision and is not in alignment with increased precision - falls by type of fall (accidental, anticipated physiological, or unanticipated physiological falls). Additionally, to present fall rates by only two age groups and gender without context appropriate structure and process measures fails to inform practice, safety, or science.
**This is not currently publicly reported. There are no publicly reported measures for POs at this time. CMS is considering using this measure in accountability applications in the next 2 years.
**No unintended consequences.
**Usability is high. My only concern is that it is not anticipated that this measure will be publicly reported. I feel strongly that this measure should be part of public reporting.
**They are not publically reported- why not? These would be helpful in a publically reported accountability program.

Criterion 5: Related and Competing Measures

Related or competing measures

- 0141 : Patient Fall Rate
- 0266 : Patient Fall

Harmonization

- The numerator for the fall measure being developed for the PACE program is closely aligned with NQF-endorsed measures 0141. They use the same definition of falls, however, the proposed measure uses a different denominator that reflects fall exposure in PACE programs as opposed to hospitals. NQF-endorsed measure 0266 is limited to ambulatory surgical centers (ASCs) and is expressed per admission rather than per day.

Pre-meeting public and member comments

-

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: PACE Participant Fall Rate
IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: 5/13/2016

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus
The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes
3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.
4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.
5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.
6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

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

☒ Health outcome: Falls

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

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

☐ Process: Click here to name the process

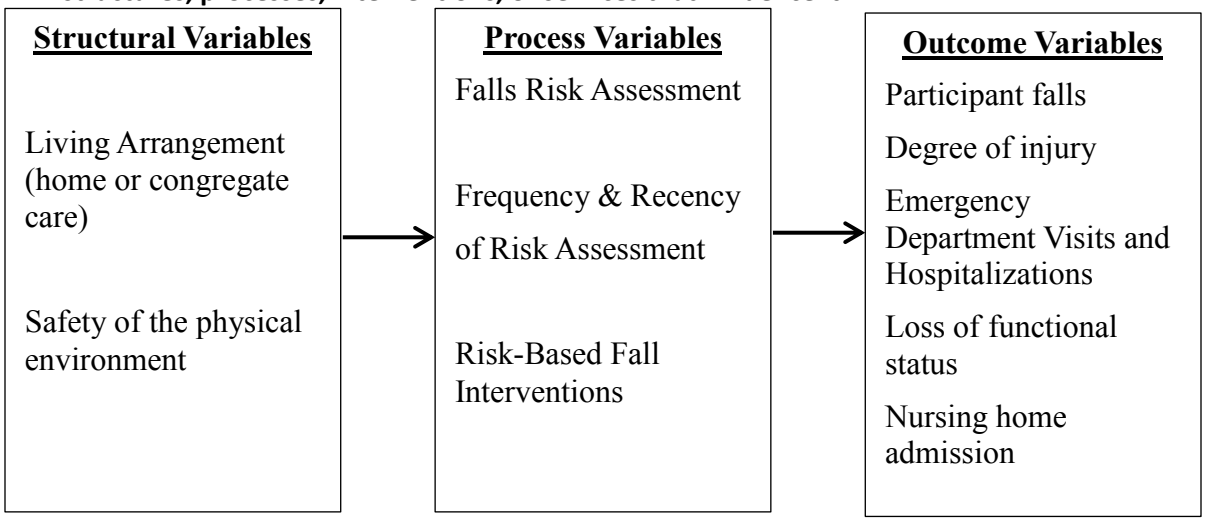
☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

_____

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.



1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

The fall rate is an individual health care outcome with structures and processes of care that can positively impact this rate. To date, there isn't published research relating falls by PACE participants with structure or process elements. Therefore, eight peer-reviewed articles on patient falls in hospitals were reviewed.

- Structural factors related to falls: These include characteristics of the nursing workforce, nurse staffing levels, Magnet status (a status awarded by the American Nurses Credentialing Center based on organization and delivery of nursing care within a health care facility), nursing turnover, and nursing work environment.
- Process factors: These include fall risk assessment, frequency of risk assessment, how recently the last risk assessment was conducted, and implementation of prevention protocols.
- Strengths: All seven studies examined patient fall rates and nursing characteristics/nurse staffing at the unit level (as opposed to the hospital level). Most studies used a conceptual framework to guide the testing of the relationships between staffing and fall rates. Most studies used nursing

care hours, nursing skill mix, fall rates, and rates of falls with injury as specified by NQF or similar to NQF.

- Weaknesses: Some studies failed to use a hierarchical model of analysis (i.e., patients and nurses nested in units and, in turn, units nested in hospitals). Some studies only examined one aspect of the nursing workforce, such as examining only staffing, rather than examining multiple aspects such as staffing, experience, education, and certification. Generally, studies were cross-sectional and observational rather than experimental. Process measures (fall risk assessment and prevention protocol implementation) associated with patient fall rates were not included in any of the studies.

Results include:

- Six studies found a significant indirect relationship between some aspect of inpatient nurse staffing and fall rates (Duffield et al., 2010; Dunton, Gajewski, Klaus, & Pierson, 2007; Dunton, Gajewski, Taunton, & Moore, 2004; Lake, Shang, Klaus, & Dunton, 2010; Potter, Barr, McSweeney, & Sledge, 2003; Whitman, Kim, Davidson, Wolf, & Wang, 2002). For example, higher total nursing hours per patient day or higher proportion of hours provided by registered nurses was related to lower fall rates.
- Two studies found that the evidence on fall prevention activities (processes) is mixed. Oliver, Hopper, and Seed (2000) found through a systematic literature review and meta-analysis that fall prevention activities may have reduced fall rates by up to 25 percent. More recently, Miake-Lye, Hempel, Ganz, and Shekelle (2013) found that fall prevention strategies reduced falls by up to 30 percent, although an optimal prevention bundle was not identified.

Citations:

Dunton, N., Gajewski, B., Klaus, S., & Pierson, B. (2007). The Relationships of Nursing Workforce Characteristics to Patient Outcomes. Online Journal of Issues in Nursing, 12(3). Retrieved from http://www.nursingworld.org/MainMenuCategories/ANAMarketplace/ANAPeriodicals/OJIN/TableofContents/Volume122007/No3Sept07/NursingWorkforceCharacteristics.aspx

Dunton, N., Gajewski, B., Taunton, R. L., & Moore, J. (2004). Nurse staffing and patient falls on acute care hospital units. Nurs Outlook, 52(1), 53-59.

Duffield, C., Diers, D., O´Brien-Pallas, L., Aisbett, C., Roche, M., King, M., et al. (2010). Nursing staffing, nursing workload, the work environment and patient outcomes. Appl Nurs Res.

Lake, E. T., Shang, J., Klaus, S., & Dunton, N. E. (2010). Patient falls: Association with hospital Magnet status and nursing unit staffing. Res Nurs Health, 33(5), 413-425.

Miake-Lye, I. M., Hempel, S., Ganz, D., & Shekelle, P. (2013). Inpatient fall prevention programs as a patient safety strategy: A systematic review**.** *Annals of Internal Medicine,* 158(5), 390–396.

Oliver, D., Hopper, A., & Seed, P. (2000). Do hospital fall preventions work? A systematic review. Journal of the American Geriatrics Society, 48(12), 1679–1689.

Potter, P., Barr, N., McSweeney, M., & Sledge, J. (2003). Identifying nurse staffing and patient outcome relationships: a guide for change in care delivery. Nurs Econ, 21(4), 158-166.

Whitman, G. R., Kim, Y., Davidson, L. J., Wolf, G. A., & Wang, S. L. (2002). The impact of staffing on patient outcomes across specialty units. J Nurs Adm, 32(12), 633-639.

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

_____

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

_____

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

_____

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

_____

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

_____

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: Click here to enter date range

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for each new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

_____

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

This document contains the information submitted by measure developers/stewards, but is organized according to NQF's measure evaluation criteria and process. The item numbers refer to those in the submission form but may be in a slightly different order here. In general, the item numbers also reference the related criteria (e.g., item 1b.1 relates to subcriterion 1b).

Brief Measure Information

NQF #: 3001
De.2. **Measure Title:** PACE Participant Fall Rate
Co.1.1. **Measure Steward:** CMS
De.3. **Brief Description of Measure:** The quarterly incidence rate of falls amongst PACE participants per 1,000 participant days.
1b.1. **Developer Rationale:** Fall Rates have been found to be an important safety concern in acute care and long-term care settings. There is evidence that falls are one of the most common adverse patient events in hospitals, and they are a source of significant injury, disability, and/or death. Nearly one-third of community-dwelling individuals over age 65 fall each year (Currie, 2008). In 2013, this accounted for nearly 2.5 million injury falls—with nearly two-thirds of this number experienced by females (CDC, 2013). Several national health care organizations—including the National Quality Strategy, the Partnership for Patients, and the CMS Hospital-Acquired Condition (HAC) Reduction Program—have identified patient falls as a patient safety concern.

Every fall carries a risk of injury. Clinicians can reduce injuries in part by reducing the risk of falling. Focusing prevention efforts solely on falls with injury is a faulty approach for improving patient safety. To some extent, falls with injury are a function of patient frailty; by contrast, the total fall rate is not influenced by differences among patients' susceptibility to injury.

Many if not most falls may result in no injury or only minor injury. Nevertheless, any fall may result in emotional distress and increased risk of falling in the future. Preventing falls among the frail elderly contributes to the maintenance of the participant's functional status and place in the community and the prevention of costs of treatment associated with falls. It is important to monitor all falls, not just falls with injury.

Citations:
CDC. (2013, December). WISQARS. Retrieved December 1, 2014 from Leading Causes of Nonfatal Injury Reports, 2001–2013: http://www.cdc.gov/injury/wisqars/nonfatal.html.

Currie, L. (2008). Fall and Injury Prevention. In R. Hughes (Ed.). Patient Safety and Quality: An Evidence-Based Handbook for Nurses. Rockville, MD: AHRQ. Retrieved November 18, 2014 from http://www.ncbi.nlm.nih.gov/books/NBK2653/.

S.4. **Numerator Statement:** Falls experienced by Participants in the PACE program during the month.
S.7. **Denominator Statement:** The denominator represents exposure of PACE participants to the risk of falling.
S.10. **Denominator Exclusions:** Exclude persons who were not enrolled as PACE participants, or who were not in their home location.

De.1. **Measure Type:** Outcome
S.23. **Data Source:** Electronic Clinical Data : Electronic Health Record, Management Data, Paper Medical Records
S.26. **Level of Analysis:** Facility

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

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

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

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

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

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

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., the benefits or improvements in quality envisioned by use of this measure)*

Fall Rates have been found to be an important safety concern in acute care and long-term care settings. There is evidence that falls are one of the most common adverse patient events in hospitals, and they are a source of significant injury, disability, and/or death. Nearly one-third of community-dwelling individuals over age 65 fall each year (Currie, 2008). In 2013, this accounted for nearly 2.5 million injury falls—with nearly two-thirds of this number experienced by females (CDC, 2013). Several national health care organizations—including the National Quality Strategy, the Partnership for Patients, and the CMS Hospital-Acquired Condition (HAC) Reduction Program—have identified patient falls as a patient safety concern.

Every fall carries a risk of injury. Clinicians can reduce injuries in part by reducing the risk of falling. Focusing prevention efforts solely on falls with injury is a faulty approach for improving patient safety. To some extent, falls with injury are a function of patient frailty; by contrast, the total fall rate is not influenced by differences among patients' susceptibility to injury.

Many if not most falls may result in no injury or only minor injury. Nevertheless, any fall may result in emotional distress and increased risk of falling in the future. Preventing falls among the frail elderly contributes to the maintenance of the participant's functional status and place in the community and the prevention of costs of treatment associated with falls. It is important to monitor all falls, not just falls with injury.

Citations:

CDC. (2013, December). WISQARS. Retrieved December 1, 2014 from Leading Causes of Nonfatal Injury Reports, 2001–2013: http://www.cdc.gov/injury/wisqars/nonfatal.html.

Currie, L. (2008). Fall and Injury Prevention. In R. Hughes (Ed.). Patient Safety and Quality: An Evidence-Based Handbook for Nurses. Rockville, MD: AHRQ. Retrieved November 18, 2014 from http://www.ncbi.nlm.nih.gov/books/NBK2653/.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

A sample of 50 sites was randomly selected out of a total of 114 PACE sites. Additionally, the oldest and two newest PACE sites were included in the sample. A total of 34 of these sites submitted data from January - March 2015 for the Fall Rate. There was one (1) large outlier with 24.79 falls per 1,000 participant days. This site has an unusually low number of participant days (121) because it had only one (1) participant in January and February 2015 and two (2) participants in March 2015. This site was excluded from the analysis and site-level descriptive statistics for total participant days and total fall rates. The table below shows the descriptive statistics requested for total participant days and total falls.

Mean, Std. Dev., Median, Min, Max
Total participant days in January-March 2015 (n=33)
15,719 13,846 13,097 2,728, 77,419
Total participant falls per 1,000 participant day (n=33)
4.27 1.53 4.44 1.88, 8.59

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.**

Most of the published evidence available is primarily from hospital-based studies. Those data do show considerable variation in patient fall rates.

- Bouldin et al. (2013) examined fall rates on medical, surgical, and medical-surgical units. Fall rates were highest on medical units (4.03 falls per 1,000 patient days (PD)) and lowest on surgical units (2.56 falls per 1,000 PD).
- He et al. (2012) identified trends in fall rates by hospital unit type. The analysis showed that fall rates remained stable or declined for most unit types between 2004 and 2009. Rates for surgical units, however, increased over time, from 2.74 falls/1,000 PD to 3.19/1,000 PD in 2008, decreasing to 2.89/1,000 PD in 2009.
- Lake et al. (2010) found that fall rates were 5 percent lower in hospitals that had achieved American Nurses Credentialing Center Magnet status than in non-Magnet hospitals.

Citations:

Bouldin, E. L., Andresen, E. M., Dunton, N. E., Simon, M., Waters, T. M., Liu, M., … Shorr, R. I. (2013). Falls among adult patients hospitalized in the United States: Prevalence and trends. Journal of Patient Safety, 9(1), 13–17.

He, J., Dunton, N., & Staggs, V. (2012). Unit-level time trends in inpatient fall rates of US hospitals. Medical Care, 50, 801–807.

Lake, E. T., Shang, J., Klaus, S., & Dunton, N. E. (2010). Patient falls: Association with hospital Magnet status and nursing unit staffing. Research in Nursing & Health, 33(5), 413–425.

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*

PACE participants are frail elderly in each site, thus they may be considered a single population. We did examine fall rates based on two demographic variables—age and gender—so that the potential for sociodemographic adjustment can be assessed.

- Age is defined as the participant age at the end of the reporting month. It is to be recorded in single years

from 55 through 89. To comply with Health Insurance Portability and Accountability Act (HIPAA requirements), all participants aged 90 and above will be top coded at 90.

- Gender is to be classified as male or female.

We examined correlations among total fall rates and PACE site characteristics. Pearson product-moment correlation coefficient, or "r", was used. Pearson's r is a measure of the strength and direction of the linear relationship between two variables. To interpret the correlations between variables, we used the following parameters: r = 0.80 or higher is a very strong relationship; r = 0.60-0.79 is a strong relationship; r = 0.40-0.59 is a moderate relationship; r = 0.20-0.39 is a weak relationship; and r < 0.19 is a very weak relationship. (Evans, 1996).

Data from the feasibility study showed that the average age of PACE participants who had a fall was 77.54 with a standard deviation of 10.20 indicating that total falls are fairly tightly distributed across age for PACE participants. Almost 70% of those who had a fall were female, reflecting the gender distribution of this population. Both PACE-site mean participant age and mean proportion of males had very weak correlations with total fall rates (r = 0.08 and r = -0.14, respectively).

Citation:

Evans, J.D. (1996). Straightforward statistics for the behavioral sciences. Pacific Grove, CA: Brooks/Cole Publishing.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.

Several studies have demonstrated a difference in falls rates for specific populations. Disparities have been identified according to age (Fhon et al, 2013; CDC, 2006), gender (Steven & Sogolow, 2005; CDC, 2006), disability (Lavedan, 2014; Ranaweera et al, 2013; Lee & Stokic, 2008), and race/ethnicity (CDC, 2006). Hospitalization for hip fractures due to falls is significantly higher for females than for males. However, fatality rates due to falls are higher for men than for women, and higher for Caucasians compared to African-Americans (CDC, 2006). Among community-dwelling older women, age-adjusted fall rates are not different between African-Americans and Caucasians. However, the authors did find racial differences for location of falls and biomechanics of falls (falling forward vs. laterally), which may explain differing fall-related fracture risk between Caucasian and African-American women (Faulkner et al., 2005).

Citations:

Centers for Disease Control (CDC; 2006). Fatalities and injuries from fall among older adults – United States, 1993-2003 and 2001-2005. Morbitity and Mortality Weekly Report, 55(45), 1221-1224.

Faulkner, K. A., Cauley, J. A., Zmuda, J. M., Landsittel, D. P., Nevitt, M. C., Newman, A. B., … Redfern, M. S. (2005). Ethnic differences in the frequency and circumstances of falling in older community-dwelling women. Journal of the American Geriatrics Society, 53(10), 1774–1779. http://doi.org/10.1111/j.1532-5415.2005.53514.x

Fhon, J. R., Rosset, I., Freitas, C. P., Silva, A. O., Santos, J. L., & Rodrigues, R. A. (2013). Prevalence of falls among frail elderly adults. Rev Saude Publica, 47(2), 266-273. doi: 10.1590/s0034-8910.2013047003468.

Lavedan Santamaria, A., Jurschik Gimenez, P., Botigue Satorra, T., Nuin Orrio, C., & Viladrosa Montoy, M. (2014). [Prevalence and associated factors of falls in community-dwelling elderly.]. Aten Primaria. doi: 10.1016/j.aprim.2014.07.012.

Lee, J. E., & Stokic, D. S. (2008). Risk factors for falls during inpatient rehabilitation. Am J Phys Med Rehabil, 87(5), 341-350; quiz 351, 422. doi: 10.1097/PHM.0b013e31816ddc01.

Ranaweera, A. D., Fonseka, P., PattiyaArachchi, A., & Siribaddana, S. H. (2013). Incidence and risk factors of falls among the elderly in the District of Colombo. Ceylon Med J, 58(3), 100-106. doi: 10.4038/cmj.v58i3.5080.

Stevens, J. A., Sogolow, E. D. (2005). Gender differences for non-fatal unintentional fall related injuries among older

adults. Injury Prevention: Journal of the International Society for Child and Adolescent Injury Prevention. 11, 115–119. doi: 10.1136/ip.2004.005835.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF;
  OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**

Affects large numbers, A leading cause of morbidity/mortality, Patient/societal consequences of poor quality

1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**

Falls pose a significant economic burden. In 2012, fall-related injuries cost the Nation more than $30 billion, with the costs expected to nearly double by 2020 (CDC, 2013a). The costs to treat an individual injured by a fall average $17,500, excluding possible legal fees (Shumway-Cook, Ciol, Hoffman, Dudgeon, Yorston, & Chan, 2009). Hospitalization costs for an injury fall exceeded $34,000 in 2012. The elderly also require longer healing times and longer treatment durations, causing subsequent losses of independence and functional capacity.

Falls are the leading cause of fatal injury for people over age 65 and the most common cause of nonfatal trauma-related hospital admissions (CDC, 2013a). Nearly one-third of community-dwelling individuals in this age group fall each year (Currie, 2008). In 2013, this accounted for nearly 2.5 million injury falls—with nearly two-thirds of this number experienced by females (CDC, 2013). Injuries from falls include fractures, traumatic brain injury, and other internal trauma. Internal injuries led to 28 percent of fall-related fatalities (CDC, 2013a). The number of nonfatal falls has increased by 34 percent in the last decade, from 1.85 million in 2004 to nearly 2.5 million in 2013 (CDC, 2013).

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

CDC. (2013a). Home and Recreational Safety. Retrieved November 13, 2014 from Costs of Falls Among Older Adults: http://www.cdc.gov/homeandrecreationalsafety/falls/fallcost.html.

CDC. (2013, December). WISQARS. Retrieved December 1, 2014 from Leading Causes of Nonfatal Injury Reports, 2001–2013: http://www.cdc.gov/injury/wisqars/nonfatal.html.

Currie, L. (2008). Fall and Injury Prevention. In R. Hughes (Ed.). Patient Safety and Quality: An Evidence-Based Handbook for Nurses. Rockville, MD: AHRQ. Retrieved November 18, 2014 from http://www.ncbi.nlm.nih.gov/books/NBK2653/.

Shumway-Cook, A., Ciol, M., Hoffman, J., Dudgeon, B., Yorston, K., & Chan, L. (2009). Falls in the Medicare population: Incidence, associated factors, and impact on health care. Physical Therapy, 89(4), 1–9.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.**
(Describe how and from whom their input was obtained.)

Not applicable.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria 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. **Cross Cutting Areas** *(check all the areas that apply):*
Safety

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.)*
None at this time.

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

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

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Not applicable.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Falls experienced by Participants in the PACE program during the month.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Monthly data aggregated to quarterly reporting periods.

S.6. **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, 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.
A PACE participant fall is a sudden, unanticipated descent in which a participant comes to rest on the floor or some other surface, person, or object.

Inclusion Criteria:
- All PACE participant falls occurring in the participants home; in assisted living facilities, if that is their usual place of residence; in the PACE center, or in the care of a PACE transportation operator.
- Participants who are assisted to the floor by a care provider (assisted fall) are to be included in the count of falls.

Exclusion Criteria:
- Participants who fall (or sink) back to a bed, chair, car seat, walker seat, or toilet are excluded in the count of falls.
- Exclude falls in the participant home by staff, visitors, family members, or others who were not PACE participants
- Exclude participants who were not in their home location. For example, exclude participants who were in an emergency room, hospitalized, in a long term care facility, in a hospice facility, in skilled nursing care, in a rehabilitation setting.

Specific data collection items and responses:
- Fall Auto No.
- Month of Fall
  - January = 1
  - February = 2
  - Etc.
- Age (at end of month):
  - Age in years if 55–89
  - Age greater >89 = 90+
  - Unknown = 99
- Gender:
  - Male = 1
  - Female = 2
  - Unknown = 99

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
The denominator represents exposure of PACE participants to the risk of falling.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk : Dual eligible beneficiaries, Populations at Risk : Individuals with multiple chronic conditions, Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
Total number of PACE participant days during the calendar month. This is calculated as the sum of the PACE site participant census for each day in the month, aggregated quarterly.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Exclude persons who were not enrolled as PACE participants, or who were not in their home location.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the*

denominator such as definitions, 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)

- Exclude persons who were not enrolled as PACE participants on the specific day of the month.
- Exclude participants who were not in their home location. For example, exclude participants who were hospitalized, in a long term care facility, in a hospice facility, in skilled nursing care, in a rehabilitation setting.
- Exclude participants who were deceased for each day after the date of death.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*

Stratification will be based on characteristics of PACE programs, including caseload size, location, region of the country and academic affiliation, and years of operation.

- Caseload size varies significantly across PACE sites. Categories of caseload size will be determined after we gather information on the size of each program and size of fluctuations over the course of a year. With just over 100 PACE programs, we anticipate having no more than 3 categories so that there is a sufficient sample size to produce reliable rates in each group.

Per the U.S. Office of Management and Budget definition:

- Location
  - Metropolitan is a county or group of contiguous counties, of which one or more has a core urban area with a population of 50,000 or more. The counties are linked by social and economic integration.
  - Micropolitan is a county or group of contiguous counties, of which one or more has an urban area with at least 10,000 persons but less than 50,000 population.
  - Non-Metropolitan is a county that is not associated with a Metropolitan or Micropolitan group of counties.
- Academic affiliation will have two categories: Yes and No. Yes indicates a site that is operated by the primary clinical site for a School of Medicine. No indicates that a site is operated by another organization.
- Years of operation for PACE programs vary widely; one program has been in operation for only a few months, while another has been in operation for more than 17 years. Years of Operation is indicated in whole years and months in a partial year. At most, three categories of "Years of Operation" will be identified in order to maintain a sufficient sample in each category to support reliable reporting.

Risk Adjustment Type:

Risk stratification will be used rather than risk adjustment. Stratification will be based on PACE site characteristics. Because PACE participants are frail elderly in each site, they may be considered a single population, not requiring risk adjustment to account for different populations across PACE sites.

Two demographic variables—age and gender—will be collected so that the potential for sociodemographic adjustment can be assessed.

- Age is defined as the participant age at the end of the reporting month. It is to be recorded in single years from 55 through 89. To comply with HIPAA requirements, all participants aged 90 and above will be top coded at 90.
- Gender is to be classified as male or female.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)

Stratification by risk category/subgroup

If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with*

measure testing under Scientific Acceptability)
Not applicable.

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Provided in response box S.15a

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
Not applicable.

S.16. **Type of score:**
Ratio
If other:

S.17. **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.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
The Fall Rate is calculated as the number of falls to PACE participants per 1,000 participant days during a calendar quarter. Data are collected monthly. The calculation steps are as follows:

1. Sum the number of falls for each of the 3 months in the quarter.
2. Multiply the numerator by 1,000. This step merely facilitates interpretation of results because it reduces leading zeros in the rate.
3. List the number of PACE site participants in the census for each day in the months included in the quarter.
4. Sum the number of participants across each day.
5. Sum the number of participant days in each month.
6. Rate calculation: (Number of falls x 1,000) / (Total number of participant days).

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
No sampling is involved in data gathering for the Fall Rate.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not applicable.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Falls and Participant Days are collected each month so that the impact of missing data can be reduced. PACE sites that fail to report data for 1 month, the same month for both the numerator and denominator, will have their

quarterly rates based on 2 months of data. PACE programs that fail to report data for 2 months out of the quarter will not have rates calculated, as a 1-month sample decreases the reliability and potentially the validity of the data to an unacceptably low level.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data : Electronic Health Record, Management Data, Paper Medical Records

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
The data collection instrument is uploaded to this application as an appendix (A.1). Data are to be collected from participant clinical records, both paper and electronic. The data sources are participant clinical records from clinicians affiliated with the PACE program, including registered nurses (RNs), physical therapists (PTs), occupational therapists (OTs), physicians (MDs and DOs), nurse practitioners (NPs), and physician assistants (PAs). If the PACE participant was in an institutional setting during the reporting period, include falls documented in the clinical records from the institution, whether a hospital, emergency room, nursing home, skilled nursing facility, rehabilitation, or some other institutional setting. Data collectors should extract fall information from clinical records in those organizations as well.

Participant Days data are to be collected from participant census data. Data collectors should record the number of PACE participants on each day in the quarter and note this information in the form presented in Table 2. Partial days count as 1 day for the purpose of this measure.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Other
If other: PACE programs provide services to participants who live in their own homes (or in home-like settings) in the community. Participants attend PACE centers regularly (e.g., 3 days per week) for a variety of activities and support services. If a participant i

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
Not applicable.

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
Falls_Testing_NQF-635987552473066799.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: PACE Participant Fall Rate
Date of Submission: 5/13/2016
Type of Measure:

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

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed

performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are

different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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.23***)	**Measure Tested with Data From:**
☒ abstracted from paper record	☒ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☒ abstracted from electronic health record	☒ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

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

1.3. What are the dates of the data used in testing? January 1, 2015 through March 31, 2015

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.26*)**	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☒ other: PACE Organization	☒ other: PACE Organization

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

A sample of 50 sites was randomly selected out of a total of 114 PACE sites. Additionally, the oldest and two newest PACE sites were included in the sample. A total of 34 of these sites submitted data from January - March 2015 for the Fall Rate, with these sites having a minimum census size of 1, and a maximum of 854 participants (median = 190). There was one (1) large outlier with 24.79 falls per 1,000 participant days. This site has an unusually low number of participant days (121) because it had only one (1) participant in January and February 2015 and two (2) participants in March 2015. This site was excluded from the analysis and site-level descriptive statistics for total participant days and total fall rates. Characteristics of the PACE Organizations which submitted data are shown in Table 1 below.

Table 1. Characteristics of PACE Organizations Participating in Testing

Category	*N (%)*
Affiliated with Academic Medical Center	
Yes	3 (8.8)
No	29 (85.3)
Unknown	1 (2.9)
Location	
Metropolitan	27 (79.4)
Micropolitan	2 (5.9)
Non-metropolitan	4 (11.8)

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*)

Data were collected on a total of 1,995 falls. Many PACE organizations operate multiple program sites. To minimize respondent burden, organizations with multiple sites were instructed to report falls data for participants at their oldest site. Participants at other sites were excluded from data reporting. Characteristics or participants with falls are shown in Table 2.

Table 2: Characteristics of Participants With Falls and Characteristics of Falls

Participant-Level Characteristic (n = 587)	Summary Statistics	
Age (years)	Mean (SD) Median (Min Max)	77.54 (10.20) 79 (56, 99)
Gender	Male Female	178 (30.32%) 409 (69.68%)
Total number of falls in March 2015	Mean (SD) Median (Min Max) 0 1 2+	1.25 (0.64) 1 (1, 6) 0 (0.00%) 483 (82.30%) 104 (17.70%)
Total number of falls with injury in March 2015	Mean (SD) Median (Min Max) 0 1 2+	0.54 (0.61) 18.5 (2, 71) 301 (51.30%) 258 (44.00%) 28 (4.70%)

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.

For content validity nine experts in academia and 24 experts who are on the Technical Expert Panel for the PACE project were invited to voluntarily review the content validity of the measure instructions. Of them, 12 experts, including 10 TEP experts and 2 academic experts, independently evaluated content validity of the total falls/falls with injury measure instructions.

For reliability testing a sample of 50 sites was randomly selected out of a total of 114 PACE sites. Additionally, the oldest and two newest PACE sites were included in the sample. A total of 34 of these sites submitted data from January - March 2015 for the Fall Rate. There was one (1) large outlier with 24.79 falls per 1,000 participant days. This site has an unusually low number of participant days (121) because it had only one (1) participant in January and February 2015 and two (2) participants in March 2015.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education,

language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

The patient level demographic variables collected were age (top coded at 90) and gender.

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*)

Site-level reliability of the measure was assessed using the signal-to-noise analysis approach. This approach was originally proposed by Adams for normally distributed data using a mixed model (Adam, 2009). The assessment of reliability for the pressure ulcer rate will use a similar approach modified for binary outcomes.

The signal-to-noise analysis, which is appropriate for the measures, quantifies the amount of variation in performance due to differences in sites (signal), as opposed to differences due to random variation within each site (noise). The signal-to-noise method results in a reliability statistic that ranges from 0 to 1 for each site. A value of 0 indicates that all variation is due to random variation, and a value of 1 indicates that all variation is due to real differences in site performance. The signal-to-noise approach for reliability assessment depends on the normality assumption for the distributions of these rates. For PACE-acquired pressure ulcer rates, the distributions are not normal. One sites with fewer than 20 participants reviewed for pressure ulcers was excluded from analysis.

Citation:

Adams J. L. (2009). The reliability of provider profiling: a tutorial. RAND Corporation, Santa Monica.

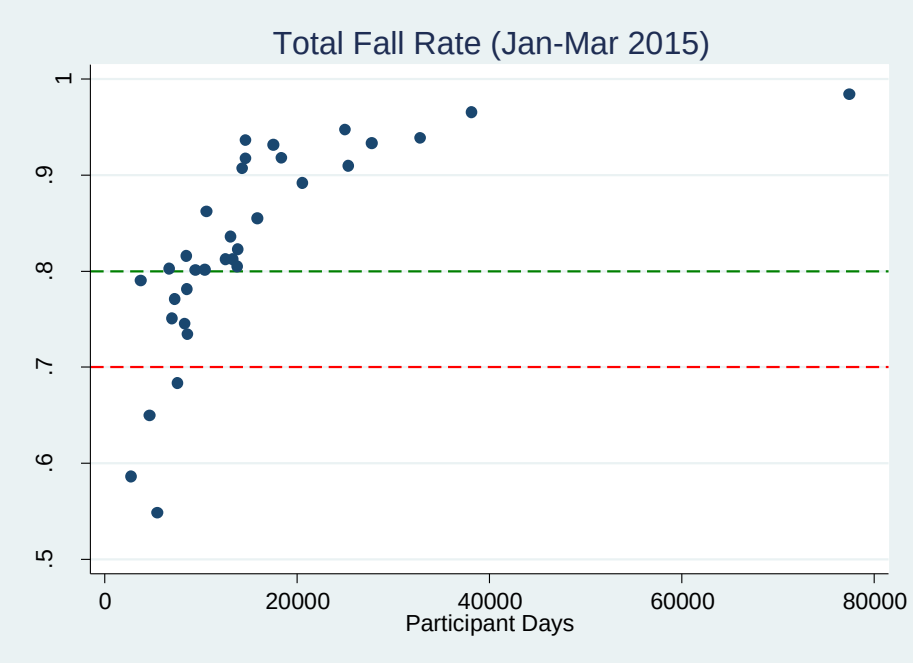
2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.g., *percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Table 3 summarizes the reliability assessment results. The mean reliability score for the Fall Rate was 0.83. When we plotted the reliability scores versus the total participant days (Figure 1), there was a highly significant direct association between the reliability score of the Total Fall Rate and total participant days (r=0.66, p<0.001).

Table 1. Signal-to-Noise Assessment of Reliability of the 3-Month Total Fall Rate

Measures	Reliability Score: Mean (SD)	Median	Minimum, Maximum
Total participant fall rate (*n*=33)	0.83 (0.10)	0.82	0.55, 0.98

Figure 1. Signal-to-Noise Reliability Scores of Total Fall Rates of 33 PACE Sites



Note: A reliability score of 0.8 or higher is considered high, and a reliability score between 0.7 and 0.8 is considered acceptable.

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 mean reliability score of the three (3)-month total participant fall rate was 0.83, higher than the cutoff value of 0.8 for high reliability, suggesting the three (3)-month fall rate is reliable in differentiating PACE sites. The median of the estimated reliability score was 0.82, the minimum was 0.55, and the maximum was 0.98. There are four (4) sites with reliability scores lower than the acceptable cutoff value of 0.7, suggesting the fall rates of these sites contained too much noise to differentiate these sites from other sites.

Reliability scores are strongly affected by the total participant days. When we plotted the reliability scores versus the total participant days (Figure 1 in section 2a2.3 above), there was a highly significant direct association between the reliability score of the Total Fall Rate and total participant days ($r=0.66$, $p<0.001$). All four (4) sites with reliability scores less than 0.7 were sites that had fewer than 8,000 total participant days. Knowing that five (5) of the 33 sites reported data for March but not for January and February, and two (2) of them had reliability scores less than 0.7, we expect reliability scores for these sites would have been higher if they provided data for all three months. We also expect the reliability scores for all would have been

higher if we collected information on falls during a longer period time (six (6) months or a year) or if more participants were enrolled in the sites.

2b2. VALIDITY TESTING

2b2.1. 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*)

2b2.2. 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)*

Content validity was assessed using a panel of 12 experts to: (1) quantify experts' degree of agreement regarding the content of the measure instructions (i.e., PACE Measure Instructions) and (2) obtain experts' narrative comments on the measure instructions. Based on the findings from content validity testing with expert review, we revised the measure. The revised instructions were distributed to PACE sample sites for pilot data collection.

Content and face validity of the measure was analyzed by calculating item-level content validity indices (I-CVIs). The I-CVI indicates the proportion of experts who consider the item as content (or face) valid. Experts rated content/face validity using a 4-point scale: 1 = very low (major modification needed), 2 = low (some modification needed), 3 = high (no modification needed but could be improved with minor changes), and 4 = very high (no modification needed). I-CVI is computed for each item by counting the number of experts giving a rating of 3 or 4 and dividing the number by the total number of experts (Polit, Beck, & Owen, 2007).

Content validity reviewed by TEP and academic experts was systematically assessed in terms applicability to PACE programs.

Citation:

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health. 30*, 459-467.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Table 4 displays I-CVIs for the Fall Rate measure.

Table 4: Content Validity Results From Experts for Data Elements in the PACE Participant Fall Rate

Fall Rate Data Element	I-CVI
Measure description	1.0 (6/6)
Definitions: • Numerator	1.0 (10/10)
• Denominator	.90 (9/10)
Measure calculation	.86 (6/7)
Exclusion criteria: • Falls by staff, visitors, or others who were not PACE participants.	1.0 (2/2)
Overall applicability of the indicator to the PACE participants and PACE sites	.92 (11/12)

Note: I-CVI = item-level content validity index; I-CVI/ave = average of I-CVIs. Each parenthesis indicates the number of experts who rated the data element as 3 or 4 divided by the total number of experts who responded.

2b2.4. 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?*)

Polit et al. (2007) suggested that items with good or acceptable content (or face) validity should have an I-CVI of .78 or higher from three or more experts' review. Based on this, we used .78 as a cutoff point to determine acceptable content (or face) validity. Another evaluation criterion was based on Lynn (1986). Lynn (1986) argued that the disagreement is accepted only if "six or more experts" gave an item a rating of 1 (very low) or 2 (low).

The findings showed acceptable content validity for the measure descriptions, definitions, measure calculations, and exclusion criteria, with the I-CVIs all greater than .78.

Overall, experts reported good content validity regarding the overall applicability of the total fall to the PACE sites and participants (I-CVI = .92).

Citations:

Lynn, M. (1986). Determination and quantification of content validity. *Nursing Research, 35,* 381–385.

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Research in Nursing & Health. 30, 459-467.

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

2b3.1. 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*)

Not tested. Excluding falls for people other than PACE participants and falls back to a bed or chair are relatively straightforward and don't warrant testing.

2b3.2. 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*)

Not applicable.

2b3.3. 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*)

Not applicable.

_____

2b4. 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 2b5.

2b4.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**
☐ **Statistical risk model with** Click here to enter number of factors **risk factors**
☒ **Stratification by 2 risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

Not applicable.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

Risk stratification was used rather than risk adjustment. Stratification was based on PACE site characteristics, such as census and years in operation. Because PACE participants are frail elderly in each site, they may be considered a single population, not requiring risk adjustment to account for different populations across PACE sites. Further data collection could result in additional risk stratification.

Two demographic variables—age and gender—were collected so that the potential for sociodemographic adjustment can be assessed.

- Age is defined as the participant age at the end of the reporting month. It is to be recorded in single years from 55 through 89. To comply with Health Insurance Portability and Accountability Act (HIPAA requirements), all participants aged 90 and above will be top coded at 90.
- Gender is to be classified as male or female.

We examined correlations among total fall rates and PACE site characteristics. Pearson product-moment correlation coefficient, or "*r*", was used. Pearson's *r* is a measure of the strength and direction of the linear relationship between two variables. To interpret the correlations between variables, we used the following parameters: $r = 0.80$ or higher is a very strong relationship; $r = 0.60$-0.79 is a strong relationship; $r = 0.40$-0.59 is a moderate relationship; $r = 0.20$-0.39 is a weak relationship; and $r < 0.19$ is a very weak relationship. (Evans, 1996).

Citation:
Evans, J.D. (1996). Straightforward statistics for the behavioral sciences. Pacific Grove, CA: Brooks/Cole Publishing.

2b4.4a. What were the statistical results of the analyses used to select risk factors?

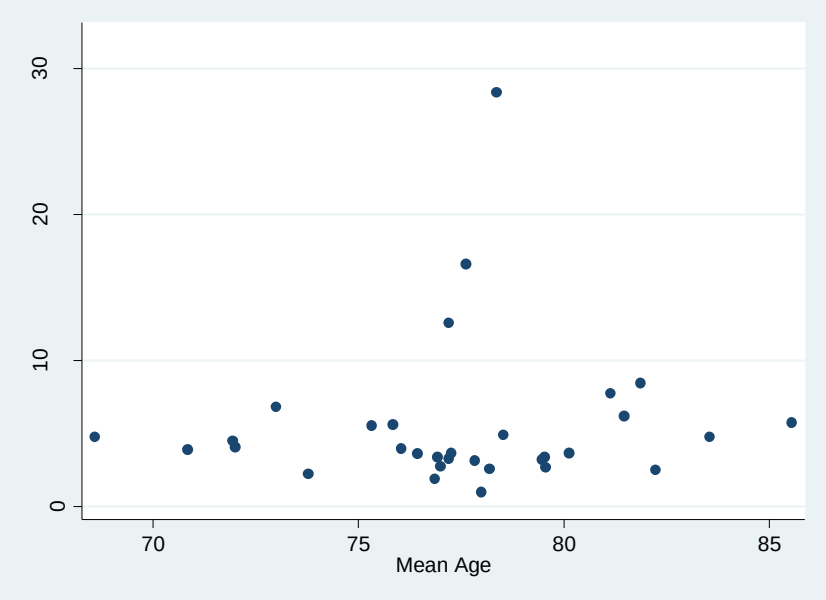
Table 2 shows that characteristics of participants with falls, and total number of falls per participant.

Table 2: Characteristics of Participants With Falls and Characteristics of Falls

Participant-Level Characteristic (*n* = 587)	Summary Statistics	
Age (years)	Mean (SD) Median (Min Max)	77.54 (10.20) 79 (56, 99)
Gender	Male Female	178 (30.32%) 409 (69.68%)
Total number of falls in March 2015	Mean (SD) Median (Min Max) 0 1 2+	1.25 (0.64) 1 (1, 6) 0 (0.00%) 483 (82.30%) 104 (17.70%)

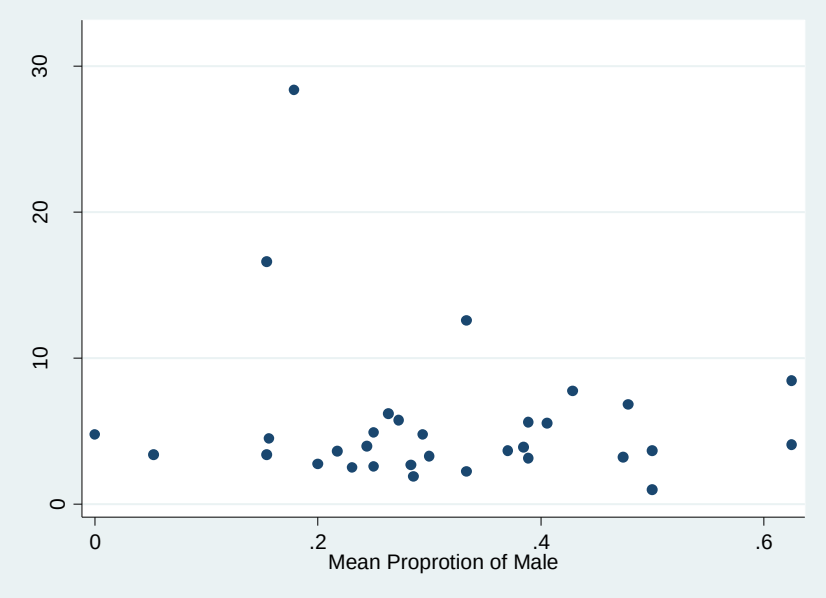
Figure 1 shows the correlations between total fall rate and the mean age of PACE site participants. Figure 2 shows the correlation between total fall rate and the mean proportion of males by PACE site. Both PACE-site participant characteristics had a very weak correlation with total fall rates.

Figure 1: Correlation Between Total Fall Rates and Mean Age (r= 0.08, n=32).
Mean age was calculated by site-level mean age of participants having falls.



Note: One PACE site excluded from scatterplot because of outlier data.

Figure 2: Correlation Between Total Fall Rates and Mean Proportion of Male (r = –0.14, n = 32).
Mean proportion of male was calculated by site-level mean proportion of males having falls. Negative correlation indicates that sites having more males with falls were likely to have lower rates of total falls.



Note: One PACE site excluded from scatterplot because of outlier data.

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

Both mean age and mean percent male were weakly correlated with fall rates (r = 0.08 and r = -0.14, respectively). After implementation of the measure, we will continue to collect data to determine the usefulness of risk-stratification based on age and gender.

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Not applicable.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

Age and sex were found to be only weakly correlated with falls rates (see Figures 1 and 2).

2b4.10. 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*)

Adequacy of sex and age to control for differences across PACE sites will be further evaluated after measure implementation with the larger population of PACE sites and participants.

2b4.11. 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*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. 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)*

Due to our small sample size, we did not conduct statistical analyses to determine differences in performance across PACE sites. However, the descriptive statistics indicate that there are

differences in total fall rates across PACE sites (mean = 6.25, SD = 6.83, median = 3.93, range = 1.0-32.06). After implementation, we will conduct further analyses to determine significant differences in performance for total fall rates across PACE sites.

2b5.2. 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*)

Not applicable.

2b5.3. 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?*)

Not applicable.

2b6. 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 SDS 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 SDS 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.***

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Not Applicable.

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

Not Applicable.

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

Not Applicable.

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

Because of the small amount of missing data, we did not conduct analyses of responders vs. nonresponders.

2b7.2. 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*)

In testing, only 1 site did not report falls or injury falls. This site had only 1 participant. Through subsequent follow-up, we determined that this 1 participant did not experience a fall or injury fall even during the data collection period. Because of the small amount of missing data, we did not conduct analyses of responders vs. nonresponders.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

Not applicable.

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), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

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)*

Some data elements are in defined fields in 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.**

Some PACE Organizations do not use electronic medical records. All organizations will abstract data manually for this measure from either their electronic or paper charts.

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.**

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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

After collecting data from PACE sites for feasibility and reliability testing, we conducted a post-data collection survey to ask PACE sites about data that they did not have available, data collection burden, and other issues. Overall, the data collection time was reasonable at 3-4 hours. While the sites reported a fairly high data collection burden, this was balanced by the fact that over half of the sites stated that the data were very easy to obtain. Further, all of the sites stated that fall rates are useful for quality improvement and 64% were supportive of

national PACE comparison data. Thus, although there is a perceived data collection burden, this is outweighed by the usefulness of the data and comparative benchmarks. Because of the high reported ease of obtaining the data, we anticipate that the perceived data collection burden will decrease as sites become more familiar with the data collection and submission process.

- Sites said that it took between 3 and 4 hours to collect the fall rate data and another hour to submit the data on-line.
- 73% of PACE sites reported that they considered the data collection burden to be medium or high burden.
- 69% of the sites reported that they collected falls data from electronic health records, although the large majority said they did manual extraction from electronic records.
- 54% of the sites said that it was very easy to obtain the data.
- 100% of responding sites said that the fall rates would be useful for quality improvement.
- 64% said that they strongly agreed with the statement that national comparison data would be helpful for quality improvement.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
None.

4. d

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.

Planned	Current Use (for current use provide URL)
Public Reporting Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Not in use	

4a.1. **For each CURRENT use, checked above, provide:**

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

Not currently in use.

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
This is a new measure. CMS is evaluating its use in upcoming PACE quality programs.

4a.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 is considering the use of the PACE Participant Fall Rate in accountability applications within the next two years.

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

Improvement data will be obtained once the measure has been implemented and tracked over time.

4b.2. **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.**
Improvement data will be obtained once the measure has been implemented and tracked over time.

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
No negative unintended consequences have been identified.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0141 : Patient Fall Rate
0266 : Patient Fall

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

5a. Harmonization

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 completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
The numerator for the fall measure being developed for the PACE program is closely aligned with NQF-endorsed measures 0141. They use the same definition of falls, however, the proposed measure uses a different denominator that reflects fall exposure in PACE programs as opposed to hospitals. NQF-endorsed measure 0266 is limited to ambulatory surgical centers (ASCs) and is expressed per admission rather than per day.

5b. Competing Measures

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

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
Not applicable.

Appendix

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

Attachment **Attachment:** AppendixA1_Falls_Data_Collection_Sheet-635987585006812369.docx

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** CMS
Co.2 **Point of Contact:** Stacy, Davis, stacy.davis@cms.hhs.gov, 410-786-7813-
Co.3 **Measure Developer if different from Measure Steward:** Econometrica, Inc.
Co.4 **Point of Contact:** Mark, Stewart, mstewart@econometricainc.com, 240-204-5168-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
The Fall Rate measure was developed in partnership with CMS by a team lead by Econometrica, Inc. consisting of Econometrica (prime contractor); the University Of Kansas Medical Center Research Institute (KUMCRI; subcontractor); Drs. Rosemary Kennedy and Barbara Resnick, and Ms. Heidi Bossley (consultants).

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:**
Ad.3 **Month and Year of most recent revision:**
Ad.4 **What is your frequency for review/update of this measure?**
Ad.5 **When is the next scheduled review/update for this measure?**

Ad.6 **Copyright statement:**
Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

A.1 Appendix PACE Participant Fall Rate

PACE Participant Fall Rate Data Abstraction Sheet

If there were no falls at your site in a month, enter 0 (zero) on the first row.

Fall No.	Month of Fall	Age (at end of month)	Gender
	1=January 2015 2=February 2015	Age in years if 55–89 90+=Age greater >89 99=Unknown	1=Male 2=Female 99=Unknown
1			
2			
3			

4			
5			
6			

Participant Census Days

January 2015	Number of Participants in Census
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 3003
Measure Title: PACE Participant Falls With Injury Rate
Measure Steward: CMS
Brief Description of Measure: The quarterly incidence rate of falls with injury amongst PACE participants per 1,000 participant days.
Developer Rationale: Fall and Falls With Injury Rates have been found to be important safety concerns in acute care and long-term care settings. There is evidence that falls are one of the most common adverse patient events in hospitals, and they are a source of significant injury, disability, and/or death. Several national health care organizations—including the National Quality Strategy, the Partnership for Patients, and the CMS Hospital-Acquired Condition (HAC) Reduction Program—have identified patient falls as a patient safety concern.

Falls With Injury may result in fatal and non-fatal injuries ranging from minor lacerations to severe head injuries (WHO, 2012). The majority of fall-related injuries are non-fatal. Several studies have demonstrated a difference in injurious fall rates for specific populations. Disparities have been identified according to age (Fhon et al., 2013) and disability, particularly cognitive impairment (Lavedan, 2014; Ranaweera et al., 2013).

Citations:

Fhon, J. R., Rosset, I., Freitas, C. P., Silva, A. O., Santos, J. L., & Rodrigues, R. A. (2013). Prevalence of falls among frail elderly adults. Rev Saude Publica, 47(2), 266–273. doi: 10.1590/s0034-8910.2013047003468

Lavedan Santamaria, A., Jurschik Gimenez, P., Botigue Satorra, T., Nuin Orrio, C., & Viladrosa Montoy, M. (2014). Prevalence and associated factors of falls in community-dwelling elderly. Aten Primaria. doi: 10.1016/j.aprim.2014.07.012

Ranaweera, A. D., Fonseka, P., Pattiya Arachchi, A., & Siribaddana, S. H. (2013). Incidence and risk factors of falls among the elderly in the District of Colombo. Ceylon Med J, 58(3), 100–106. doi: 10.4038/cmj.v58i3.5080

World Health Organization. (2012). Falls. Retreived from http://www.who.int/mediacentre/factsheets/fs344/en/index.html.

Numerator Statement: Falls with injury experienced by participants in the PACE program during the

month.
Denominator Statement: The denominator represents exposure of PACE participants to the risk of falling.
Denominator Exclusions: Exclude persons who were not enrolled as PACE participants, or who were not in their home location.

Measure Type: Outcome
Data Source: Electronic Clinical Data : Electronic Health Record, Management Data, Paper Medical Records
Level of Analysis: Facility

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

Summary of evidence:

- The falls with injury rate is an individual health care outcome. The fall rate is an individual health care outcome with structures and processes of care that can positively impact this rate. To date, there isn't published research relating falls by PACE participants with structure or process elements.
- The developers reviewed eight peer-reviewed articles on patient falls in hospitals and summarized the strengths and weaknesses of those studies. Overall, these studies found a significant indirect relationship between some aspect of inpatient nursing staffing and fall rates. Two studies found the evidence on fall prevention activities (processes) is mixed. One study found through a systematic literature review and meta-analysis that fall prevention activities may have reduced fall rates by up to 25 percent. Another study found that fall prevention strategies reduced falls up to 30 percent , although an optimal prevention bundle was not identified.

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*

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

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

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

- The developers collected data from a sample of 50 sites that were randomly selected out of a total of 114 PACE sites. A total of 34 sites submitted data from January –March 2015 for fall rate. One site was excluded.

- The developers found a 1.78 mean participant falls with injury rate (n=33). They concluded that there are performance gaps in falls with injury and cited a study that reported falls with injury rates in acute inpatient units varied by unit type and over time.

Disparities:

- The developers examined falls with injury rates based on two demographic variables, age and gender, so that the potential for sociodemographic adjustment could be assessed. The found that falls with injury are fairly tightly distributed across age for PACE participants and both PACE-site mean participant age and mean proportion of makes had very weak/negligible correlation with injury fall rates. Several studies have demonstrated a difference in fall rates for specific populations. Disparities have been identified in age, gender, and race/ethnicity.

Questions for the Committee:

- *Specific question on information provided for gap in care.*
- *Is there a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*

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

1c. Composite - Quality Construct and Rationale
Maintenance measures – same emphasis on quality construct and rationale as for new measures.

Committee pre-evaluation comments

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

1a. Evidence to Support Measure Focus

Comments: **The relationship between the fall with injury has a diagram of structure and process variables. The variables are not related to fall injury as an outcome. The process measures are listed as fall risk assessment (not fall injury risk assessment of history of falls with injury); frequency and recency of risk assessment – the timeframe is not specified; risk-based fall
Structure of care are living arrangements (home or congregate care) and safety of the physical environment

No mentions are made about the PACE program: such as the composition of the teams and implementation of the program – integrity of the model and team for health care and difficulty of the program – PACE is neither a healthcare provider or a healthcare plan (Hirth, V., Baskins,. J., and Dever-Bumba, M. (2009). Program of all-inclusive care (PACE): Past, Present and Future. JAMDA. Mar. 2009. …..PACE began in 1973…

Also, the literature review does not support the patient safety measure: fall with injury. Most of the literature is dated, and focused on falls – more focused on hospital. Since PACE is a program designed to keep patients in their homes and communities and reduce hospitals and nursing home admissions, it is unclear why the literature review did not focus on community falls and injury. There is no discussion of risk by age (85 and older) or injury risk factors (osteoporosis and anticoagulation) that should be assessment upon elderly admission to PACE:

Significant literature review missing is:
Fhon article does not address injury – this was an epidemiological, cross sectional study that enrolled 240 elderly from Riberio Preto, Sao Palo state; collected data from Nov 2010 – Feb 2011: found increased fall prevalence in women with younger age (60-79), with 28% experience 1-2 fall; 84.7% of the falls were in bathrooms, 55.9% lost balance, and 54.2% had scratches.

Berry. S.D., and Miller, FM. (2008). Falls: Epidemiology, Pathophysiology, and Relationship to Fracture Curr Osteoporos Rep. 2008 December ; 6(4): 149–154. "Approximately 30% of falls result in an injury that requires medical attention and with fractures occurring in approximately 10% of falls. Fractures associated with falls are multi-factorial in origin. In addition to the traditional risk factors for falls, the fall descent, fall impact, and bone strength are all important determinants of whether a fracture will occur as a result of an event. "
van den Berg M, Castellote J, Mahillo-Fernandez I, Pedro-Cuesta J: Incidence
of spinal cord injury worldwide: a systematic review. Neuroepidemiology 2010, 34:184–192. Reported falls were leading cause of SCI in the elderly
Bimodal distribution: 15-29 years; and >= 65 yoa

Oliver, et al., (2010). Preventing falls and fall-related injuries in hospitals. Clinics in Geriatric Medicine, Nov.

Stevens, J., Mahoney, J.E., Ehrenreich, H. (2014). Circumstances and outcomes of falls among high risk community-dwelling older adults. Injury Epidemiology 2014, 1:5. Examined outcomes of falls in the Wisconsin SAFE program (Safety Assessment for Elders) experienced by 328 patients: Data were available for 1,172 falls. A generalized linear mixed model analysis showed that being aged ≥ 85 (OR = 2.1, 95% confidence interval [CI] = 1.2‑3.9), female (OR = 2.1, 95% CI = 1.3‑3.4), falling backward and landing flat (OR = 5.6, 95% CI = 2.9– 10.5), sideways (OR = 4.6, 95% CI = 2.6– 8.0) and forward (OR = 3.3, 95% CI = 2.0– 5.7) were significantly associated with the likelihood of injury. Of 783 falls inside the home, falls in the bathroom were more than twice as likely to result in an injury compared to falls in the living room (OR = 2.4, 95% CI = 1.2– 4.9).

Oliver, D., Healey, F., & Haines, T. P. (2010). Preventing falls and fall-related injuries in hospitals 26(4): 645-92. Clinical Geriatric Medicine 26(4), 645-692. 30-51% of hospital falls result in injury.

Cameron, I.D., Gillespie, L.D., Robertson, M,C,, Murray, G.R., Hill, K.D., Cumming, R.G., & Kerse, N. (2012). Interventions for preventing falls in older people in care facilities and hospitals. Cochrane Database of Systematic Reviews. 12: CD005465. DOI: 10.1002/14651858.CD005465.pub3.

Levant, S., Chari, K., & DeFrances, C.J. (2015). Hospitalizations for patients age 85 and over in the United States, 2000-2010. NCHS Data Brief. No. 182. Available at: hppt://www.cdc.gov/nchs/data/databriefs/db182.htm.

**The evidence is well know about the burden and consequences of falls in the aging population - 65 and older. The PACE organization enrolls participants who meet their criteria starting at age 55, which is outside of the evidence. But their target population is divided into two age groups - for which there is no explanation. This reference should be made to the following original research:

Levant, S., Chari, K., & DeFrances, C.J. (2015). Hospitalizations for patients age 85 and over in the United States, 2000-2010. NCHS Data Brief. No. 182. Available at: hppt://www.cdc.gov/nchs/data/databriefs/db182.htm.

**This is an outcome measure. The developer used 8 peer-reviewed articles on patient falls in hospitals which found a significant indirect relationship between some aspect of nursing care and fall rates. Fall rates were shown to be positively affected by fall prevention strategies in 2 studies, but no optimal prevention bundle was identified. Literature on falls in PACE does not exist. As insurer and provider, PACE organizations serve participants at home, in the PACE center, in AL, SNF and acute care, among other settings.

**The fall rate is an individual health care outcome with structures and processes of care that can positively impact this rate (diagram included). Eight peer-reviewed articles on patient falls in hospitals were reviewed (no articles on PACE). There is significant indirect relationship between some aspect of inpatient nursing staffing and fall rates. Fall prevention activities may have reduced fall rates by up to 25 percent.

1b. Performance Gap

Comments: **The PACE Participant is considered as a "single population" which resulted in their decision to

examine falls and injurious falls by only age and gender. Only two age groups are defined (55-89) and 90 and older, male/female. This is over aggregated and unable PACE to hazard risk adjust or examine gaps based on increased vulnerability (known fallers, history of hip fracture, anticoagulation, head trauma, SCI, etc). However, a feasibility study was report in the application (1.b.4) that reports average age of participants with fall and injury was 76.88 (SD 10.33) and 72% of those who fell where female. This supports the need to focus practice on injury risk assessment and injury history and injury prevention.
**34 sites submitted data for Jan-March 2015. A 1.78 mean participant fall rate was found. The developers concluded that there are performance gaps, citing a study that reported falls with injury rates in hospitals vary by unit type and over time.
**The performance data provided demonstrate a gap in care. The developers collected data from 33 sites randomly selected out of 114 PACE sites, from January to March 2015 for fall rate. They found a 1.78 mean participant falls with injury rate for PACE, compared to 50th percentile of 0.96 for medical unit, 0.88 for surgical unit, and 1.21 for medical/surgical unit.

Criteria 2: Scientific Acceptability of Measure Properties
2a. Reliability **2a1. Reliability Specifications** **Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures**
2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented. **Data source(s):** Numerator: A PACE participant fall with injury is a sudden, unanticipated descent in which a participant comes to rest on the floor or some other surface, person, or object, resulting in an injury level of minor or greater. Injury Level: Injury levels should be assessed 24 hours after the fall and be categorized as: • None: Participant had no injuries (no signs of symptoms) resulting from the fall; if an x ray, CT scan, or other post fall evaluation results in a finding of no injury. • Minor: Resulted in application of dressing, cleaning wound, ice, limb evaluation, topical medication, pain, bruise, or abrasion. • Moderate: Resulted in wound treatment such as suturing, skin glue, steri-strips, or splint; possible muscle or joint strain. • Major: Resulted in fracture, surgery, casting, traction, or required neurological or internal injury consultation. Possibly resulting in hospitalization or in permanent loss of function. • Death: Participant died as a result of injuries from the fall. Inclusion Criteria: • All PACE participant falls with injury occurring in the participants home; in assisted living facilities, if that is their usual place of residence; in the PACE center, or in the care of a PACE transportation operator.

- Participants who are injured when assisted to the floor by a care provider (assisted fall) are to be included in the count of falls with injury.

Exclusion Criteria:

- Participants who fall (or sink) back to a bed, chair, car seat, walker seat, or toilet are excluded in the count of falls with injury.
- Exclude falls in the participant home by staff, visitors, family members, or others who were not PACE participants
- Exclude participants who were not in their home location. For example, exclude participants who were in an emergency room, hospitalized, in a long term care facility, in a hospice facility, in skilled nursing care, in a rehabilitation setting.

Specific data collection items and responses:

- Fall Auto No.
- Month of Fall
  - January = 1
  - February = 2
  - Etc.
- Age (at end of month):
  - Age in years if 55–89
  - Age greater >89 = 90+
  - Unknown = 99
- Gender:
  - Male = 1
  - Female = 2
  - Unknown = 99
- Injury Level
  - None = 1
  - Minor = 2
  - Moderate = 3
  - Major = 4
    - Death = 5

Denominator:

The denominator represents exposure of PACE participants to the risk of falling.

Questions for the Committee :

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*

Is it likely this measure can be consistently implemented?

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review: N/A

A sample of 50 sites was randomly selected out of a total of 114 PACE sites. Additionally, the oldest and two newest PACE sites were included in the sample. A total of 34 of these sites submitted data from January - March 2015 for the Falls With Injury Rate, with these sites having a minimum census size of 1, and a maximum of 854 participants (median = 190). One (1) site had only one (1) participant in January and February 2015 and two (2) participants in March 2015. The enrollment for this site was too low to provide a reliable Falls With Injury rate.

Data were collected on a total of 876 falls with injury. Many PACE organizations operate multiple program sites. To minimize respondent burden, organizations with multiple sites were instructed to report data for participants at their oldest site.

SUMMARY OF TESTING

Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

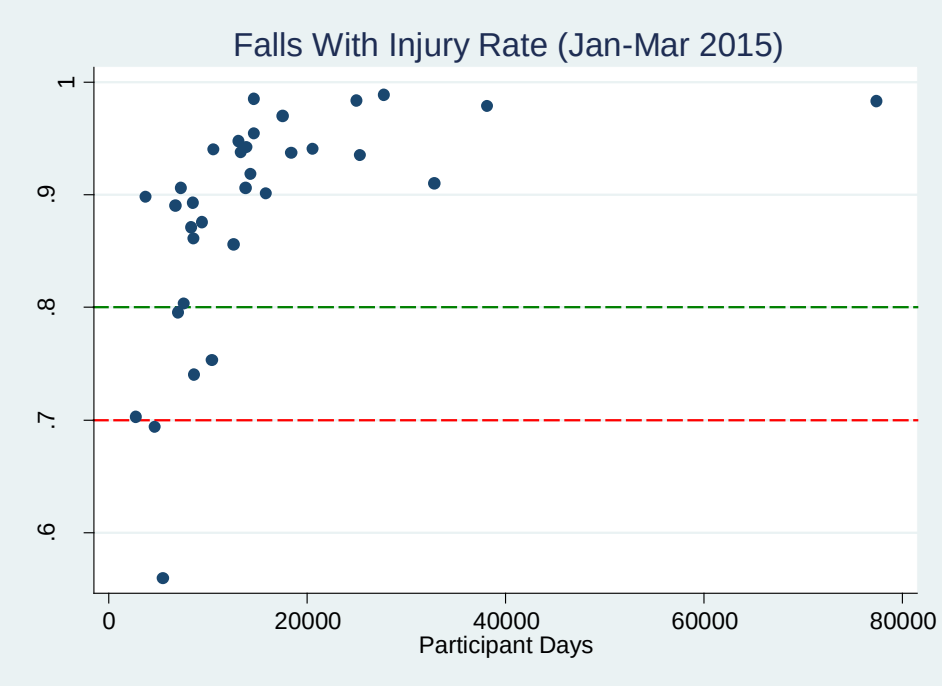
Method(s) of reliability testing Signal-to-noise analysis

Results of reliability testing [Results of reliability testing]

Table 3. Signal-to-Noise Assessment of Reliability of the 3-Month Falls With Injury Rate

Measures	Reliability Score: Mean (SD)	Median	Minimum, Maximum
Total participant falls with injury rate (*n*=33)	0.88 (0.10)	0.91	0.56, 0.99

Figure 1. Signal-to-Noise Reliability Scores of Falls With Injury Rates of 33 PACE Sites



Note: A reliability score of 0.8 or higher is considered high, and a reliability score between
0.7 and 0.8 is considered acceptable.

Guidance from the Reliability Algorithm Based on this, since the developers provided statistical testing and the reliability was high 0.88 on average, the reliability for this measure is "HIGH."

Questions for the Committee:

- *Specific questions on the method and results of reliability testing.*
- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

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

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

○ *Are the specifications consistent with the evidence?*

2b2. Validity testing

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 content validity nine experts in academia and 24 experts who are on the Technical Expert Panel for the PACE project were invited to voluntarily review the content validity of the measure instructions. Of them, 12 experts, including 10 TEP experts and 2 academic experts, independently evaluated content validity of the total falls/falls with injury measure instructions.

SUMMARY OF TESTING

Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

- ☒ **Face validity only**
- ☐ **Empirical validity testing of the measure score**

Validity testing method: Content validity was assessed using a panel of 12 experts to: (1) quantify experts' degree of agreement regarding the content of the measure instructions (i.e., PACE Measure Instructions) and (2) obtain experts' narrative comments on the measure instructions. Based on the findings from content validity testing with expert review, we revised the measure. The revised instructions were distributed to PACE sample sites for pilot data collection.

Validity testing results:

Table 4 displays I-CVIs for the Falls With Injury Rate measure.

Table 4: Content Validity Results for Data Elements in the Total Falls/Falls With Injury Measure Instructions

Total Falls/Falls With Injury Data Element	I-CVI
Measure description	1.0 (6/6)
Definitions:	
• Numerator	1.0 (10/10)
• Denominator	.90 (9/10)
• Injury level:	
o None	1.0 (9/9)
o Minor	
o Moderate	
o Major	

Measure calculation .86 (6/7)
Exclusion criteria:

- Falls by staff, visitors, or others who were not PACE participants.1.0 (2/2)

Overall applicability of the indicator to the PACE participants and PACE sites .92 (11/12)
Note: I-CVI = item-level content validity index; Each parenthesis indicates the number of experts who rated the data element as 3 or 4 divided by the total number of experts who responded.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*
- *Other specific question of the validity testing?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

The developer wrote, "Not tested. Excluding falls for people other than PACE participants and falls back to a bed or chair are relatively straightforward and don't warrant testing."

Questions for the Committee:

- *Is this a sufficient response?*

2b4. Risk adjustment: **Risk-adjustment method** ☐ **None** ☐ **Statistical model** ☒ **Stratification**

Risk-stratification by 2 risk categories: Age and Gender.

Conceptual rationale for SDS factors included ? ☐ **Yes** ☒ **No**

SDS factors included in risk model? ☐ **Yes** ☒ **No**

Risk adjustment summary

The developer wrote, "Both mean age and mean percent male were very weakly correlated with falls with injury rates (r = 0.02 and r = -0.07, respectively). After implementation of the measure, we will continue to collect data to determine the usefulness of risk-stratification based on age and gender."

Questions for the Committee:

- *Is there sufficient evidence provided by the developer to justify risk stratification by these metrics?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

The developer wrote," Due to our small sample size, we did not conduct statistical analyses to determine differences in performance across PACE sites. However, the descriptive statistics indicate that there are differences in falls with injury rates across PACE sites (mean = 2.21, SD = 2.52, median = 1.47, range = 0.21-10.48). After implementation, we will conduct further analyses to determine significant differences in performance for falls with injury rates across PACE sites."

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

N/A

2b7. Missing Data
The developer wrote, "Because of the small amount of missing data, we did not conduct analyses of responders vs. non-responders."

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

Committee pre-evaluation comments

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

2a1. & 2b1. Specifications

Comments: **The fall with injury definition is consistent, but to aggregate the fall with injury to the entire PACE Organization is inconsistent. Fall and injury rates vary by setting of care: LTC (MDS), Home Health Service (OASIS), Outpatient Clinics (NCQA, Hospitals (NDNQI)). To have comparison of fall and injury rates analyzed and compared with PACE organizations control and compare for structure and processes.
**Specifications are clear. Exclusions of falls that occur in SNFs, EDs, hospitals, etc are part of this measure because these falls are counted by those particular settings in accordance with other measures, and there is a concern about double counting.
**The data elements are defined with specific inclusion and exclusion criteria.

2a2. Reliability Testing
Comments: **Signal to noise testing was completed. Results indicate 0.88 on average - high reliability.
**The process question to answer is: what percent of participants have a completed physical assessment of the presence and severity of injury within 24 hours of a fall (2.a.1)

**Measure score reliability testing was done using signal-to-noise analysis. The reliability score on average was 0.88 = high. This was based on 33 PACE sites over a 3 month period for a total of 876 falls with injuries.

**Reliability testing is done with signal-to-noise analysis for 3-month falls with injury rate from 33 PACE sites (286 patients total). The developers provided statistical testing and the reliability was high at 0.88 on average.

2b2. Validity Testing
Comments: **Face Validity was determined by a panel of 12 experts to quantify experts' degree of agreement regarding the content of the measure structures (two age groups, two gender, severity of injury levels, total number of fall with injury); obtained experts' narrative comments on the measure instructions.

The Validity was not tested. Excluding falls for people other than PACE participants and falls back to a bed or chair are relatively straight forward and don't warrant testing. This decision should be questioned, considering 60% of falls occur in the home, 30% in the community, and 10% in hospitals. The falls with injury are limited to those who have a fall w/ injury in their home, assisted living facility, the PACE center, or in the care of a PACE transportation operator. Each is a very different context that must be examined for validity to be determined. Face validity only by expert panel of 12.

10/10 for the numerator
9/10 for the denominator

**Measure score validity testing was done using face validity. The TEP of 12 experts was used to assess agreement with the content of the measure instructions and to obtain narrative comments. Overall applicability

of the indicator scored .92
**The measure is an indicator of quality based on experience and the literature in other settings serving a similar patient population.

**Content validity was assessed using a panel of 12 experts to quantify their degree of agreement regarding PACE Measure Instructions. There is high agreement reported.

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **Evidence is not cited to justify the number and rationale for exclusion from this quality/safety outcome measure. The exclusion criteria states: exclude participants who were not in their home location - this is insistent with falls that occur in the PACE organization - or the Type 11 Fall with Injury QI.
**Threats to validity are also reported to be that both mean age and mean percent male were weakly correlated with falls with injury rates (r=0.02 and r=-0.07). After the measure is implemented, they will conduct further analysis,
**Because of their small sample size, they did not statistical analyses across to determine differences across PACE its; but yet provide mean fall rates with injury. They will conduct further analyses after implementation of the measure.

**Exclusions of falls for non-PACE participants makes sense. Falls back into bed, chair are also excluded.
The developer will continue to collect data during implementation to determine the usefulness of risk-stratification by age and gender.

**Threats to validity was not tested since excluding falls for people other than PACE participants and falls back to a bed or chair are relatively straightforward. Age and gender differences were weakly correlated with falls with injury rates (could be due to small sample size).

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.

- This measure is generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score) Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)
- Some data elements are in defined fields in a combination of electronic sources.
- Some PACE Organizations do not use electronic medical records. All organizations will abstract data manually for this measure from either their electronic or paper charts.
- After collecting data from PACE sites for feasibility and reliability testing, a post-data collection survey was conducted, to ask PACE sites about data that they did not have available, data collection burden, and other issues.
- Some sites reported a fairly high data collection burden, however, this was balanced by the fact that over half of the sites stated that the data were very easy to obtain. Although there is a

perceived data collection burden, this is outweighed by the usefulness of the data and comparative benchmarks.
- Because of the high reported ease of obtaining the data, we anticipate that the perceived data collection burden will decrease as sites become more familiar with the data collection and submission process.
- No fees or licensing requirements to use any aspect of the measure as specified, were reported.

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

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **The burden of data collection is addressed. Many PACE organizations organizations, and during reliability testing to report data from the participants in the oldest site to reduce respondent burden. Data are collected from EMR and paper records. Personnel, while providing care, collect the data. During feasibility and reliability testing, over half the sites considered burden to be fairly high; if the data for injury severity is to be collected, the question is how severity can be determine only during provision of care - extent of injury is not always determined at the time, or 1 or 2 days after a fall.

**As noted in the measure 3001, there is a fairly high data collection burden for PACE sites because of the need to do most data collection manually for this measure. However, the data are easy to obtain from the electronic and paper records.

**In a post data collection survey, the PACE sites reported data collection time of 3-4 hours. 73% of PACE sites rated medium or high data collection burden (requires manual abstraction), and 54% reported data is fairly easy to obtain. All respondents said that the fall rates would be useful for quality improvement.

Criterion 4: Usability and Use

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

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

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

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

OR

Planned use in an accountability program? ☒ Yes ☐ No

Accountability program details:

- This is a new measure and is not currently in use.
- CMS is considering the use of the PACE Participant Fall Rate in accountability applications within the next two years

Improvement results:

- Improvement data will be obtained once the measure has been implemented and tracked over time.

Unexpected findings (positive or negative) during implementation:

- No negative unintended consequences have been identified.

Potential harms:

- No unexpected findings reported.

Feedback :

- Developer did not identify any specific feedback loops related to this 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

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **As developed, the importance of this quality and safety measure lacks precision, the variable of age is too aggregated and there is no control or consideration of injury risk of history. Other programs have more precise data for fall injury: OASIS, MDS, NDNQI. Some participants in the PACE program may also be receiving care from home health, which has Fall and Injury Data Measures, along with structure and processes. It is unknown to this reviewer if PACE receives program evaluation and quality improvement data from OASIS for their participants at the site, state, or national level.

**The measure is not currently publicly reported and is not used in an accountability program. CMS is considering using the measure in accountability applications in the next 2 years.

**This is a new measure and is not currently in use, but CMS is considering use for accountability within the next

two years.

Criterion 5: Related and Competing Measures
Related or competing measures • 0202 : Falls with injury • 0674 : Percent of Residents Experiencing One or More Falls with Major Injury (Long Stay) **Harmonization** • The numerator for the falls with injury measure being developed for the PACE program is closely aligned with NQF-endorsed measures 0202. They use the same description of injury levels, however, the proposed measure uses a different denominator that reflect fall exposure in PACE programs as opposed to hospitals. NQF-endorsed measure 0266 is limited to long-stay nursing facility residents with major injuries from falls rather than any injury.

Pre-meeting public and member comments

Submitted by: QUA INC

Strongly suggest that this measure includes data re the urgency of the task, i.e., whether patients chose to walk to the bathroom rather than wait for lift, personal assistance, etc. See this reference for inpatient setting:
http://www.patientsafetysolutions.com/docs/December_22_2009_Falls_on_Toileting_Activities.htmhttp://www.patientsafetysolutions.com/docs/December_22_2009_Falls_on_Toileting_Activities.htm

Literature supports multifactorial nature of falls, sensitive to the medications, changes in hemodynaic function. Not aware of studies reporting the frequency distribution of the tasks associated with a fall, importance of innovative design of assistive equipment design to support self-care to avoid situations as outlined in recent NYT article:

http://www.nytimes.com/2016/07/21/nyregion/insurance-groups-in-new-york-improperly-cut-home-care-hours.html.

Capture the intersection of patient and staff safety, interact with safe patient handling community at www.asphp.org for more information.

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

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

Measure Title: PACE Participant Falls With Injury Rate

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

Date of Submission: 5/13/2016

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

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

Outcome

☒ Health outcome: Falls

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

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

☐ Process: Click here to name the process

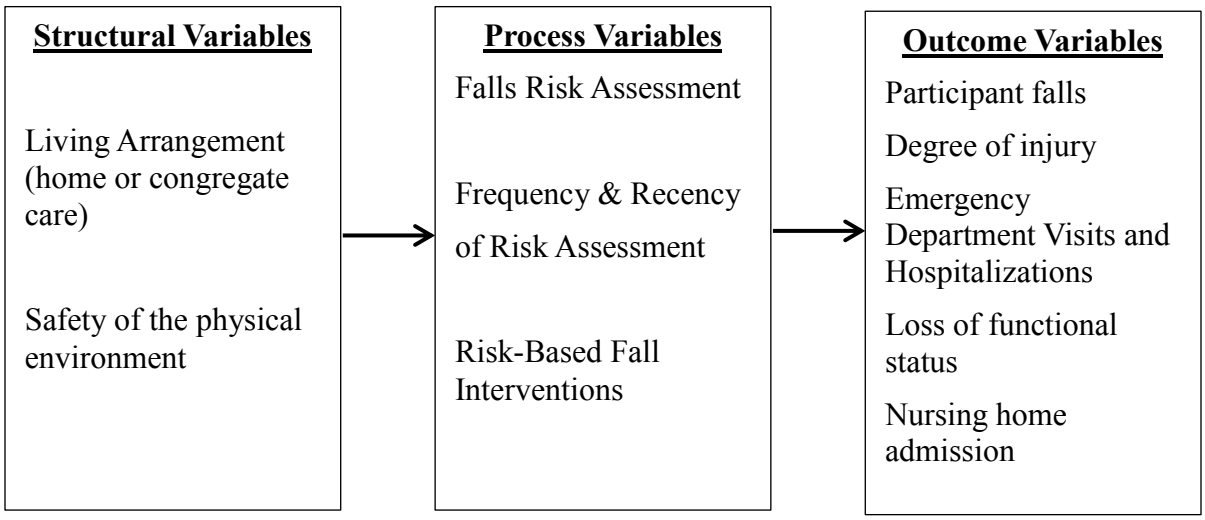
☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

_____

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.



1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

The falls with injury rate is an individual health care outcome. The fall rate is an individual health care outcome with structures and processes of care that can positively impact this rate. To date, there isn't published research relating falls by PACE participants with structure or process elements. Therefore, eight peer-reviewed articles on patient falls in hospitals were reviewed.

- Structural factors related to falls: These include characteristics of the nursing workforce, nurse staffing levels, Magnet status (a status awarded by the American Nurses Credentialing Center based on organization and delivery of nursing care within a health care facility), nursing turnover, and nursing work environment.
- Process factors: These include fall risk assessment, frequency of risk assessment, how recently the last risk assessment was conducted, and implementation of prevention protocols.
- Strengths: All seven studies examined patient fall rates and nursing characteristics/nurse staffing at the unit level (as opposed to the hospital level). Most studies used a conceptual framework to guide the testing of the relationships between staffing and fall rates. Most studies used nursing

care hours, nursing skill mix, fall rates, and rates of falls with injury as specified by NQF or similar to NQF.

- Weaknesses: Some studies failed to use a hierarchical model of analysis (i.e., patients and nurses nested in units and, in turn, units nested in hospitals). Some studies only examined one aspect of the nursing workforce, such as examining only staffing, rather than examining multiple aspects such as staffing, experience, education, and certification. Generally, studies were cross-sectional and observational rather than experimental. Process measures (fall risk assessment and prevention protocol implementation) associated with patient fall rates were not included in any of the studies.

Results include:

- Six studies found a significant indirect relationship between some aspect of inpatient nurse staffing and fall rates (Duffield et al., 2010; Dunton, Gajewski, Klaus, & Pierson, 2007; Dunton, Gajewski, Taunton, & Moore, 2004; Lake, Shang, Klaus, & Dunton, 2010; Potter, Barr, McSweeney, & Sledge, 2003; Whitman, Kim, Davidson, Wolf, & Wang, 2002). For example, higher total nursing hours per patient day or higher proportion of hours provided by registered nurses was related to lower fall rates.
- Two studies found that the evidence on fall prevention activities (processes) is mixed. Oliver, Hopper, and Seed (2000) found through a systematic literature review and meta-analysis that fall prevention activities may have reduced fall rates by up to 25 percent. More recently, Miake-Lye, Hempel, Ganz, and Shekelle (2013) found that fall prevention strategies reduced falls by up to 30 percent, although an optimal prevention bundle was not identified.

Citations:

Dunton, N., Gajewski, B., Klaus, S., & Pierson, B. (2007). The Relationships of Nursing Workforce Characteristics to Patient Outcomes. Online Journal of Issues in Nursing, 12(3). Retrieved from http://www.nursingworld.org/MainMenuCategories/ANAMarketplace/ANAPeriodicals/OJIN/TableofContents/Volume122007/No3Sept07/NursingWorkforceCharacteristics.aspx

Dunton, N., Gajewski, B., Taunton, R. L., & Moore, J. (2004). Nurse staffing and patient falls on acute care hospital units. Nurs Outlook, 52(1), 53-59.

Duffield, C., Diers, D., O´Brien-Pallas, L., Aisbett, C., Roche, M., King, M., et al. (2010). Nursing staffing, nursing workload, the work environment and patient outcomes. Appl Nurs Res.

Lake, E. T., Shang, J., Klaus, S., & Dunton, N. E. (2010). Patient falls: Association with hospital Magnet status and nursing unit staffing. Res Nurs Health, 33(5), 413-425.

Miake-Lye, I. M., Hempel, S., Ganz, D., & Shekelle, P. (2013). Inpatient fall prevention programs as a patient safety strategy: A systematic review. Annals of Internal Medicine, 158(5), 390–396.

Oliver, D., Hopper, A., & Seed, P. (2000). Do hospital fall preventions work? A systematic review. Journal of the American Geriatrics Society, 48(12), 1679–1689.

Potter, P., Barr, N., McSweeney, M., & Sledge, J. (2003). Identifying nurse staffing and patient outcome relationships: a guide for change in care delivery. Nurs Econ, 21(4), 158-166.

Whitman, G. R., Kim, Y., Davidson, L. J., Wolf, G. A., & Wang, S. L. (2002). The impact of staffing on patient outcomes across specialty units. J Nurs Adm, 32(12), 633-639.

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

_____

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

_____

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

_____

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

_____

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

_____

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: Click here to enter date range

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

_____

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

NATIONAL QUALITY FORUM

Measure Information

This document contains the information submitted by measure developers/stewards, but is organized according to NQF's measure evaluation criteria and process. The item numbers refer to those in the submission form but may be in a slightly different order here. In general, the item numbers also reference the related criteria (e.g., item 1b.1 relates to subcriterion 1b).

Brief Measure Information

NQF #: 3003
De.2. **Measure Title:** PACE Participant Falls With Injury Rate
Co.1.1. **Measure Steward:** CMS
De.3. **Brief Description of Measure:** The quarterly incidence rate of falls with injury amongst PACE participants per 1,000 participant days.
1b.1. **Developer Rationale:** Fall and Falls With Injury Rates have been found to be important safety concerns in acute care and long-term care settings. There is evidence that falls are one of the most common adverse patient events in hospitals, and they are a source of significant injury, disability, and/or death. Several national health care organizations—including the National Quality Strategy, the Partnership for Patients, and the CMS Hospital-Acquired Condition (HAC) Reduction Program—have identified patient falls as a patient safety concern.

Falls With Injury may result in fatal and non-fatal injuries ranging from minor lacerations to severe head injuries (WHO, 2012). The majority of fall-related injuries are non-fatal. Several studies have demonstrated a difference in injurious fall rates for specific populations. Disparities have been identified according to age (Fhon et al., 2013) and disability, particularly cognitive impairment (Lavedan, 2014; Ranaweera et al., 2013).

Citations:
Fhon, J. R., Rosset, I., Freitas, C. P., Silva, A. O., Santos, J. L., & Rodrigues, R. A. (2013). Prevalence of falls among frail elderly adults. Rev Saude Publica, 47(2), 266–273. doi: 10.1590/s0034-8910.2013047003468

Lavedan Santamaria, A., Jurschik Gimenez, P., Botigue Satorra, T., Nuin Orrio, C., & Viladrosa Montoy, M. (2014). Prevalence and associated factors of falls in community-dwelling elderly. Aten Primaria. doi: 10.1016/j.aprim.2014.07.012

Ranaweera, A. D., Fonseka, P., Pattiya Arachchi, A., & Siribaddana, S. H. (2013). Incidence and risk factors of falls among the elderly in the District of Colombo. Ceylon Med J, 58(3), 100–106. doi: 10.4038/cmj.v58i3.5080

World Health Organization. (2012). Falls. Retreived from http://www.who.int/mediacentre/factsheets/fs344/en/index.html.

S.4. **Numerator Statement:** Falls with injury experienced by participants in the PACE program during the month.
S.7. **Denominator Statement:** The denominator represents exposure of PACE participants to the risk of falling.
S.10. **Denominator Exclusions:** Exclude persons who were not enrolled as PACE participants, or who were not in their home location.

De.1. **Measure Type:** Outcome

S.23. **Data Source:** Electronic Clinical Data : Electronic Health Record, Management Data, Paper Medical Records
S.26. **Level of Analysis:** Facility

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

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

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

De.4. **IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?** The Falls With Injury Rate is not paired or grouped.

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

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

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., the benefits or improvements in quality envisioned by use of this measure)*

Fall and Falls With Injury Rates have been found to be important safety concerns in acute care and long-term care settings. There is evidence that falls are one of the most common adverse patient events in hospitals, and they are a source of significant injury, disability, and/or death. Several national health care organizations—including the National Quality Strategy, the Partnership for Patients, and the CMS Hospital-Acquired Condition (HAC) Reduction Program—have identified patient falls as a patient safety concern.

Falls With Injury may result in fatal and non-fatal injuries ranging from minor lacerations to severe head injuries (WHO, 2012). The majority of fall-related injuries are non-fatal. Several studies have demonstrated a difference in injurious fall rates for specific populations. Disparities have been identified according to age (Fhon et al., 2013) and disability, particularly cognitive impairment (Lavedan, 2014; Ranaweera et al., 2013).

Citations:

Fhon, J. R., Rosset, I., Freitas, C. P., Silva, A. O., Santos, J. L., & Rodrigues, R. A. (2013). Prevalence of falls among frail elderly adults. Rev Saude Publica, 47(2), 266–273. doi: 10.1590/s0034-8910.2013047003468

Lavedan Santamaria, A., Jurschik Gimenez, P., Botigue Satorra, T., Nuin Orrio, C., & Viladrosa Montoy, M. (2014). Prevalence and associated factors of falls in community-dwelling elderly. Aten Primaria. doi: 10.1016/j.aprim.2014.07.012

Ranaweera, A. D., Fonseka, P., Pattiya Arachchi, A., & Siribaddana, S. H. (2013). Incidence and risk factors of falls among the elderly in the District of Colombo. Ceylon Med J, 58(3), 100–106. doi: 10.4038/cmj.v58i3.5080

World Health Organization. (2012). Falls. Retreived from http://www.who.int/mediacentre/factsheets/fs344/en/index.html.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

A sample of 50 sites was randomly selected out of a total of 114 PACE sites. Additionally, the oldest and two newest PACE sites were included in the sample. A total of 34 of these sites submitted data from January - March 2015 for the Fall Rate. There was one (1) large outlier with 24.79 falls per 1,000 participant days. This site has an unusually low number of participant days (121) because it had only one (1) participant in January and February 2015 and two (2) participants in March 2015. This site was excluded from the analysis and site-level descriptive statistics for total participant days and total fall rates. The table below shows the distributions of total participant days and total fall rates.

	Mean,	Std. Dev.,	Median,	Min,	Max
Total participant days in January–March 2015 (n=33)					
	15,719	13,846	13,097	2,728	77,419
Total participant Falls With Injury per 1,000 participant day (n=33)					
	1.78	1.21	1.58	0.27,	5.48

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.

There are performance gaps in the falls with injury rate. Bouldin et al. (2013) report that falls with injury rates in acute inpatient units varied by unit type and over time (see table below).

In a study using data from the National Database of Nursing Quality Indicators®, falls with injury were measured if patients had an injury level of minor or greater. He et al. (2012) found the same results. All unit types experienced decreases in fall rates between 2004 and 2009, except for surgical units. Surgical units experienced an increase in fall rates over the period.

Falls With Injury Rate, 2008. Falls With Injury × 1,000/Total Patient Days

Unit Type		Percentiles				
		10th	25th	50th	75th	90th
Medical		0.26	0.59	0.96	1.36	1.79
Surgical	0.08	0.31	0.57	0.88	1.24	
Medical/Surgical	0.17	0.49	0.83	1.21	1.36	

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*

PACE participants are frail elderly in each site, thus they may be considered a single population. We did examine falls with injury rates based on two demographic variables—age and gender—so that the potential for sociodemographic adjustment can be assessed.

- Age is defined as the participant age at the end of the reporting month. It is to be recorded in single years from 55 through 89. To comply with Health Insurance Portability and Accountability Act (HIPAA requirements), all participants aged 90 and above will be top coded at 90.
- Gender is to be classified as male or female.

We examined correlations among falls with injury rates and PACE site characteristics. Pearson product-moment correlation coefficient, or "r", was used. Pearson's r is a measure of the strength and direction of the linear relationship between two variables. To interpret the correlations between variables, we used the following parameters: r = 0.80 or higher is a very strong relationship; r = 0.60-0.79 is a strong relationship; r = 0.40-0.59 is a moderate relationship; r = 0.20-0.39 is a weak relationship; and r < 0.19 is a very weak relationship. (Evans, 1996).

Data from the feasibility study showed that the average age of PACE participants who fell and had an injury was 76.88 with a standard deviation of 10.33 indicating that falls with injury are fairly tightly distributed across age for PACE participants. Seventy-two percent (72%) of those who had an injurious fall were female, reflecting the gender distribution of this population. Both PACE-site mean participant age and mean proportion of males had very weak/negligible correlation with injury fall rates (r = 0.02 and r = -0.07, respectively).
Citation:
Evans, J.D. (1996). Straightforward statistics for the behavioral sciences. Pacific Grove, CA: Brooks/Cole Publishing.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.
Several studies have demonstrated a difference in falls rates for specific populations. Disparities have been identified according to age (Fhon et al, 2013; CDC, 2006), gender (Steven & Sogolow, 2005; CDC, 2006), disability (Lavedan, 2014; Ranaweera et al, 2013; Lee & Stokic, 2008), and race/ethnicity (CDC, 2006). Hospitalization for hip fractures due to falls is significantly higher for females than for males. However, fatality rates due to falls are higher for men than for women, and higher for Caucasians compared to African-Americans (CDC, 2006). Among community-dwelling older women, age-adjusted fall rates are not different between African-Americans and Caucasians. However, the authors did find racial differences for location of falls and biomechanics of falls (falling forward vs. laterally), which may explain differing fall-related fracture risk between Caucasian and African-American women (Faulkner et al., 2005).

Citations:
Centers for Disease Control (CDC; 2006). Fatalities and injuries from fall among older adults – United States, 1993-2003 and 2001-2005. Morbitity and Mortality Weekly Report, 55(45), 1221-1224.

Faulkner, K. A., Cauley, J. A., Zmuda, J. M., Landsittel, D. P., Nevitt, M. C., Newman, A. B., … Redfern, M. S. (2005). Ethnic differences in the frequency and circumstances of falling in older community-dwelling women. Journal of the American Geriatrics Society, 53(10), 1774–1779. http://doi.org/10.1111/j.1532-5415.2005.53514.x
Fhon, J. R., Rosset, I., Freitas, C. P., Silva, A. O., Santos, J. L., & Rodrigues, R. A. (2013). Prevalence of falls among frail elderly adults. Rev Saude Publica, 47(2), 266-273. doi: 10.1590/s0034-8910.2013047003468.

Lavedan Santamaria, A., Jurschik Gimenez, P., Botigue Satorra, T., Nuin Orrio, C., & Viladrosa Montoy, M. (2014). [Prevalence and associated factors of falls in community-dwelling elderly.]. Aten Primaria. doi: 10.1016/j.aprim.2014.07.012.

Lee, J. E., & Stokic, D. S. (2008). Risk factors for falls during inpatient rehabilitation. Am J Phys Med Rehabil, 87(5), 341-350; quiz 351, 422. doi: 10.1097/PHM.0b013e31816ddc01.

Ranaweera, A. D., Fonseka, P., PattiyaArachchi, A., & Siribaddana, S. H. (2013). Incidence and risk factors of falls among the elderly in the District of Colombo. Ceylon Med J, 58(3), 100-106. doi: 10.4038/cmj.v58i3.5080.

Stevens, J. A., Sogolow, E. D. (2005). Gender differences for non-fatal unintentional fall related injuries among older adults. Injury Prevention: Journal of the International Society for Child and Adolescent Injury Prevention. 11, 115–119. doi: 10.1136/ip.2004.005835.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF;
  OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**

Affects large numbers, A leading cause of morbidity/mortality, Patient/societal consequences of poor quality

1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**

Falls pose a significant economic burden. In 2012, fall-related injuries cost the Nation more than $30 billion, with the costs expected to nearly double by 2020 (CDC, 2013a). The costs to treat an individual injured by a fall average $17,500, excluding possible legal fees (Shumway-Cook, Ciol, Hoffman, Dudgeon, Yorston, & Chan, 2009). Hospitalization costs for an injury fall exceeded $34,000 in 2012. The elderly also require longer healing times and longer treatment durations, causing subsequent losses of independence and functional capacity.

Falls are the leading cause of fatal injury for people over age 65 and the most common cause of nonfatal trauma-related hospital admissions (CDC, 2013a). Nearly one-third of community-dwelling individuals in this age group fall each year (Currie, 2008). In 2013, this accounted for nearly 2.5 million injury falls—with nearly two-thirds of this number experienced by females (CDC, 2013). Injuries from falls include fractures, traumatic brain injury, and other internal trauma. Internal injuries led to 28 percent of fall-related fatalities (CDC, 2013a). The number of nonfatal falls has increased by 34 percent in the last decade, from 1.85 million in 2004 to nearly 2.5 million in 2013 (CDC, 2013).

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

CDC. (2013a). Home and Recreational Safety. Retrieved November 13, 2014 from Costs of Falls Among Older Adults: http://www.cdc.gov/homeandrecreationalsafety/falls/fallcost.html.

CDC. (2013, December). WISQARS. Retrieved December 1, 2014 from Leading Causes of Nonfatal Injury Reports, 2001–2013: http://www.cdc.gov/injury/wisqars/nonfatal.html.

Currie, L. (2008). Fall and Injury Prevention. In R. Hughes (Ed.). Patient Safety and Quality: An Evidence-Based Handbook for Nurses. Rockville, MD: AHRQ. Retrieved November 18, 2014 from http://www.ncbi.nlm.nih.gov/books/NBK2653/.

Shumway-Cook, A., Ciol, M., Hoffman, J., Dudgeon, B., Yorston, K., & Chan, L. (2009). Falls in the Medicare population: Incidence, associated factors, and impact on health care. Physical Therapy, 89(4), 1–9.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

Not applicable.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria 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. **Cross Cutting Areas** *(check all the areas that apply):*
Safety

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.)*
None at this time.

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

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

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Not applicable. New measure.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Falls with injury experienced by participants in the PACE program during the month.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Monthly data aggregated to quarterly reporting periods.

S.6. **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, 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.

A PACE participant fall with injury is a sudden, unanticipated descent in which a participant comes to rest on the floor or some other surface, person, or object, resulting in an injury level of minor or greater.

Injury Level: Injury levels should be assessed 24 hours after the fall and be categorized as:

- None: Participant had no injuries (no signs of symptoms) resulting from the fall; if an x ray, CT scan, or other post fall evaluation results in a finding of no injury.
- Minor: Resulted in application of dressing, cleaning wound, ice, limb evaluation, topical medication, pain, bruise, or abrasion.
- Moderate: Resulted in wound treatment such as suturing, skin glue, steri-strips, or splint; possible muscle or joint strain.
- Major: Resulted in fracture, surgery, casting, traction, or required neurological or internal injury consultation. Possibly resulting in hospitalization or in permanent loss of function.
- Death: Participant died as a result of injuries from the fall.

Inclusion Criteria:

- All PACE participant falls with injury occurring in the participants home; in assisted living facilities, if that is their usual place of residence; in the PACE center, or in the care of a PACE transportation operator.
- Participants who are injured when assisted to the floor by a care provider (assisted fall) are to be included in the count of falls with injury.

Exclusion Criteria:

- Participants who fall (or sink) back to a bed, chair, car seat, walker seat, or toilet are excluded in the count of falls with injury.
- Exclude falls in the participant home by staff, visitors, family members, or others who were not PACE participants
- Exclude participants who were not in their home location. For example, exclude participants who were in an emergency room, hospitalized, in a long term care facility, in a hospice facility, in skilled nursing care, in a rehabilitation setting.

Specific data collection items and responses:

- Fall Auto No.
- Month of Fall
  - January = 1
  - February = 2
  - Etc.
- Age (at end of month):
  - Age in years if 55–89
  - Age greater >89 = 90+
  - Unknown = 99
- Gender:
  - Male = 1
  - Female = 2
  - Unknown = 99
- Injury Level
  - None = 1
  - Minor = 2
  - Moderate = 3
  - Major = 4
  - Death = 5
  - Unknown = 99

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
The denominator represents exposure of PACE participants to the risk of falling.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk : Dual eligible beneficiaries, Populations at Risk : Individuals with multiple chronic conditions, Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
Total number of PACE participant days during the calendar month. This is calculated as the sum of the PACE site participant census for each day in the month, aggregated quarterly.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Exclude persons who were not enrolled as PACE participants, or who were not in their home location.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
- Exclude persons who were not enrolled as PACE participants on the specific day of the month.
- Exclude participants who were not in their home location. For example, exclude participants who were hospitalized, in a long term care facility, in a hospice facility, in skilled nursing care, in a rehabilitation setting.
- Exclude participants who were deceased for each day after the date of death.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
Stratification will be based on characteristics of PACE programs, including caseload size, location, region of the country and academic affiliation, and years of operation.
- Caseload size varies significantly across PACE sites. Categories of caseload size will be determined after we gather information on the size of each program and size of fluctuations over the course of a year. With just over 100 PACE programs, we anticipate having no more than 3 categories so that there is a sufficient sample size to produce reliable rates in each group.

Per the U.S. Office of Management and Budget definition:
- Location
  - Metropolitan is a county or group of contiguous counties, of which one or more has a core urban area with a population of 50,000 or more. The counties are linked by social and economic integration.
  - Micropolitan is a county or group of contiguous counties, of which one or more has an urban area with at least 10,000 persons but less than 50,000 population.
  - Non-Metropolitan is a county that is not associated with a Metropolitan or Micropolitan group of counties.
- Academic affiliation will have two categories: Yes and No. Yes indicates a site that is operated by the primary clinical site for a School of Medicine. No indicates that a site is operated by another organization.
- Years of operation for PACE programs vary widely; one program has been in operation for only a few months, while another has been in operation for more than 17 years. Years of Operation is indicated in whole years and months in a partial year. At most, three categories of “Years of Operation” will be identified in order to maintain a sufficient sample in each category to support reliable reporting.

Risk Adjustment Type:

Risk stratification will be used rather than risk adjustment. Stratification will be based on PACE site characteristics. Because PACE participants are frail elderly in each site, they may be considered a single population, not requiring risk adjustment to account for different populations across PACE sites.

Two demographic variables—age and gender—will be collected so that the potential for sociodemographic adjustment can be assessed.

- Age is defined as the participant age at the end of the reporting month. It is to be recorded in single years from 55 through 89. To comply with HIPAA requirements, all participants aged 90 and above will be top coded at 90.
- Gender is to be classified as male or female.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Stratification by risk category/subgroup
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Not applicable.

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Provided in response box S.15a

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
Not applicable.

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **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.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
The Falls With Injury Rate is calculated as the number of falls with injury to PACE participants per 1,000 participant days during a calendar quarter. Data are collected monthly and reported quarterly. The calculation steps are as follows:

1. Sum the number of falls with injury for each of the 3 months in the quarter.
2. Multiply the numerator by 1,000. This step merely facilitates interpretation of results because it reduces leading zeros in the rate.
3. List the number of PACE site census for each day for each of the months included in the quarter.
4. Sum the number of participants across each day.
5. Sum the number of participant days in each month.
6. Rate calculation: (Number of Falls With Injury x 1,000) / (Total number of participant days

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
No sampling is involved in data gathering for the Falls With Injury Rate

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not applicable.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Falls With Injury and Participant Days are collected by month so that the impact of missing data can be reduced. PACE sites that fail to report data for 1 month, the same month for both the numerator and denominator, will have their quarterly rates based on 2 months of data. PACE programs that fail to report data for 2 months out of the quarter will not have rates calculated, as a 1-month sample decreases the reliability and potentially the validity of the data to an unacceptably low level.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data : Electronic Health Record, Management Data, Paper Medical Records

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
The data collection instrument is uploaded as an appendix (A.1) to this application. Data are to be collected from participant clinical records, both paper and electronic. The data sources are participant clinical records from clinicians affiliated with the PACE program, including RNs, PTs, OTs, physicians (MDs and DOs), NPs, and PAs.

Participant Days data are to be collected from participant census data. Data collectors should record the number of PACE participants on each day in the quarter and record this information in the form presented in the appendix. Partial days count as 1 day for the purpose of this measure.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Other
If other: PACE programs provide services to participants who live in their own homes (or in home-like settings) in the community. Participants attend PACE centers regularly (e.g., 3 days per week) for a variety of activities and support services. If a participant i

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
Not applicable.

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
FallsWithInjury_Testing_NQF.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: PACE Participant Falls With Injury Rate
Date of Submission: 5/13/2016
Type of Measure:

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

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed

performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are

different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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.23***)	**Measure Tested with Data From:**
☒ abstracted from paper record	☒ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☒ abstracted from electronic health record	☒ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

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

1.3. What are the dates of the data used in testing? Click here to enter date range

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.26*)**	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☒ other: PACE Organization	☒ other: PACE Organization

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

A sample of 50 sites was randomly selected out of a total of 114 PACE sites. Additionally, the oldest and two newest PACE sites were included in the sample. A total of 34 of these sites submitted data from January - March 2015 for the Falls With Injury Rate, with these sites having a minimum census size of 1, and a maximum of 854 participants (median = 190). One (1) site had only one (1) participant in January and February 2015 and two (2) participants in March 2015. The enrollment for this site was too low to provide a reliable Falls With Injury rate. Characteristics of the PACE Organizations which submitted data are shown in Table 1 below.

Table 1. Characteristics of PACE Organizations Participating in Testing

Category	*N (%)*
Affiliated with Academic Medical Center	
Yes	3 (8.8)
No	29 (85.3)
Unknown	1 (2.9)
Location	
Metropolitan	27 (79.4)
Micropolitan	2 (5.9)
Non-metropolitan	4 (11.8)

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*)

Data were collected on a total of 876 falls with injury. Many PACE organizations operate multiple program sites. To minimize respondent burden, organizations with multiple sites were instructed to report data for participants at their oldest site. Participants at other sites were excluded from data reporting. Characteristics of participants with injurious falls are shown in Table 2.

Table 2: Characteristics of Participants With Injurious Falls and Characteristics of Injurious Falls

Participant-Level Characteristic (n=286)	Summary Statistics	
Age (years)	Mean (SD) Median (Min Max)	76.88 (10.33) 78 (56, 99)
Gender	Male Female	81 (28.32%) 205 (71.68%)
Total number of falls with injury in March 2015	Mean (SD) Median (Min Max) 1 2+	1.11 (0.37) 1 (1, 4) 258 (90.21%) 28 (0.79%)

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.

For content validity nine experts in academia and 24 experts who are on the Technical Expert Panel for the PACE project were invited to voluntarily review the content validity of the measure instructions. Of them, 12 experts, including 10 TEP experts and 2 academic experts, independently evaluated content validity of the total falls/falls with injury measure instructions.

For reliability testing a sample of 50 sites was randomly selected out of a total of 114 PACE sites. Additionally, the oldest and two newest PACE sites were included in the sample. A total of 34 of these sites submitted data from January - March 2015 for the Falls With Injury Rate. There was one (1) large outlier with 24.79 falls per 1,000 participant days. This site has an unusually low number of participant days (121) because it had only one (1) participant in January and February 2015 and two (2) participants in March 2015.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

The patient level demographic variables collected were age (top coded at 90) and gender.

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*)

Site-level reliability of the measure was assessed using the signal-to-noise analysis approach. This approach was originally proposed by Adams for normally distributed data using a mixed model (Adam, 2009). The assessment of reliability for the pressure ulcer rate will use a similar approach modified for binary outcomes.

The signal-to-noise analysis, which is appropriate for the measures, quantifies the amount of variation in performance due to differences in sites (signal), as opposed to differences due to random variation within each site (noise). The signal-to-noise method results in a reliability statistic that ranges from 0 to 1 for each site. A value of 0 indicates that all variation is due to random variation, and a value of 1 indicates that all variation is due to real differences in site performance. The signal-to-noise approach for reliability assessment depends on the normality assumption for the distributions of these rates. For PACE-acquired pressure ulcer rates, the distributions are not normal. One sites with fewer than 20 participants reviewed for pressure ulcers was excluded from analysis.

Citation:

Adams J. L. (2009). The reliability of provider profiling: a tutorial. RAND Corporation, Santa Monica.

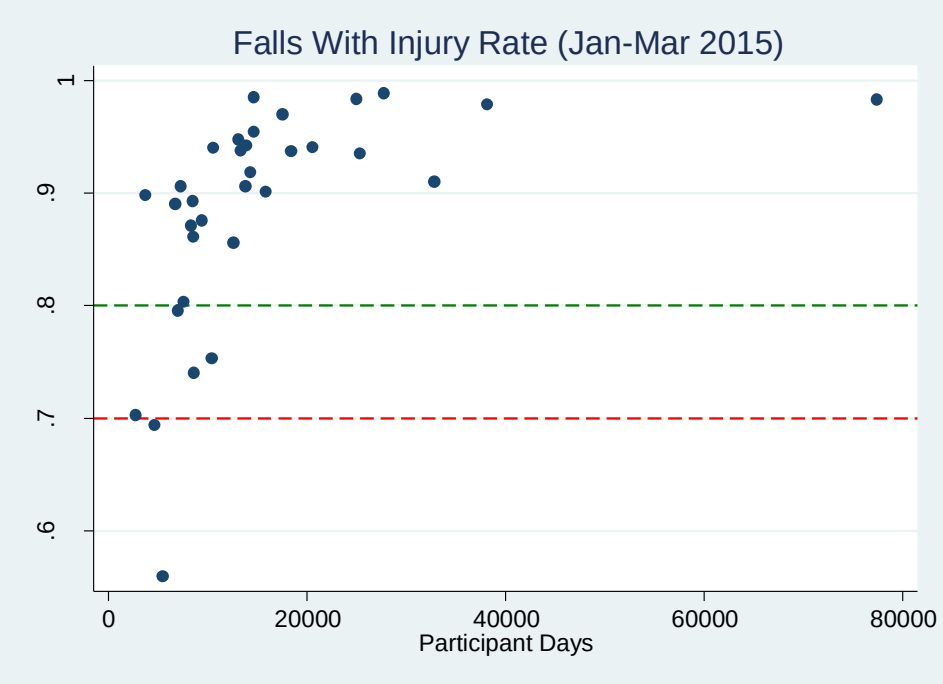
2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Table 3 summarizes the reliability assessment results. The mean reliability score for the Falls With Injury Rate was 0.88. When we plotted the reliability scores versus the total participant days (Figure 1), there was a significant direct association between the reliability score of the total participant fall rate and total participant days (r=0.50, p=0.03).

Table 3. Signal-to-Noise Assessment of Reliability of the 3-Month Falls With Injury Rate

Measures	Reliability Score: Mean (SD)	Median	Minimum, Maximum
Total participant falls with injury rate (*n*=33)	0.88 (0.10)	0.91	0.56, 0.99

Figure 1. Signal-to-Noise Reliability Scores of Falls With Injury Rates of 33 PACE Sites



Note: A reliability score of 0.8 or higher is considered high, and a reliability score between 0.7 and 0.8 is considered acceptable.

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 mean reliability score of the three (3)-month total participant fall with injury rate was 0.88, much higher than the cutoff value of 0.8 for high reliability, suggesting that the three (3)-month Falls With Injury rate was reliable in differentiating sites. The median of the estimated reliability score was 0.91, the minimum is 0.56, and the maximum was 0.99. There were two (2) sites with reliability scores lower than the acceptable cutoff value of 0.7, suggesting the Falls With Injury rates of these sites contained too much noise to differentiate them from other sites.

Reliability scores are strongly affected by the total participant days. When we plotted the reliability scores versus the total participant days (Figure 1 in section 2a2.3 above), there was a significant direct association between the reliability score of the total participant fall rate and total participant days (r=0.50, p=0.03). Knowing that five (5) of the 33 sites reported data at Round One but not at Round Two and one (1) of them has a reliability score less than 0.7, we expect reliability scores for these sites to

have been higher if they had provided data during Round Two as well. We also expect the reliability scores for all sites to be higher if we collect information on Falls With Injury during a longer period time (six (6) months or a year) or if more participants were enrolled in these sites.

2b2. VALIDITY TESTING

2b2.1. 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*)

2b2.2. 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)*

Content validity was assessed using a panel of 12 experts to: (1) quantify experts' degree of agreement regarding the content of the measure instructions (i.e., PACE Measure Instructions) and (2) obtain experts' narrative comments on the measure instructions. Based on the findings from content validity testing with expert review, we revised the measure. The revised instructions were distributed to PACE sample sites for pilot data collection.

Content and face validity of the measure was analyzed by calculating item-level content validity indices (I-CVIs). The I-CVI indicates the proportion of experts who consider the item as content (or face) valid. Experts rated content/face validity using a 4-point scale: 1 = very low (major modification needed), 2 = low (some modification needed), 3 = high (no modification needed but could be improved with minor changes), and 4 = very high (no modification needed). I-CVI is computed for each item by counting the number of experts giving a rating of 3 or 4 and dividing the number by the total number of experts (Polit, Beck, & Owen, 2007).

Content validity reviewed by TEP and academic experts was systematically assessed in terms applicability to PACE programs.

Citation:

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health. 30*, 459-467.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Table 4 displays I-CVIs for the Falls With Injury Rate measure.

Table 4: Table Error! No text of specified style in document.**-1: Content Validity Results for Data Elements in the Total Falls/Falls With Injury Measure Instructions**

Total Falls/Falls With Injury Data Element	I-CVI
Measure description	1.0 (6/6)
Definitions: • Numerator	1.0 (10/10)
• Denominator	.90 (9/10)
• Injury level: ○ None	1.0 (9/9)
○ Minor	
○ Moderate	
○ Major	
Measure calculation	.86 (6/7)
Exclusion criteria: • Falls by staff, visitors, or others who were not PACE participants.	1.0 (2/2)
Overall applicability of the indicator to the PACE participants and PACE sites	**.92 (11/12)**

Note: I-CVI = item-level content validity index; Each parenthesis indicates the number of experts who rated the data element as 3 or 4 divided by the total number of experts who responded.

2b2.4. 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?*)

Polit et al. (2007) suggested that items with good or acceptable content (or face) validity should have an I-CVI of .78 or higher from three or more experts' review. Based on this, we used .78 as a cutoff point to determine acceptable content (or face) validity. Another evaluation criterion was based on Lynn (1986). Lynn (1986) argued that the disagreement is accepted only if "six or more experts" gave an item a rating of 1 (very low) or 2 (low).

The findings showed acceptable content validity for the measure descriptions, definitions, measure calculations, and exclusion criteria, with I-CVIs all greater than .78.

Overall, experts reported good content validity regarding the overall applicability of the falls with injury measure to the PACE sites and participants (I-CVI = .92).

Citations:

Lynn, M. (1986). Determination and quantification of content validity. *Nursing Research, 35,* 381–385.

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Research in Nursing & Health. 30, 459-467.

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

2b3.1. 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*)

Not tested. Excluding falls for people other than PACE participants and falls back to a bed or chair are relatively straightforward and don't warrant testing.

2b3.2. 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*)

Not applicable.

2b3.3. 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*)

Not applicable.

2b4. 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 2b5.

2b4.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**

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

☒ **Stratification by 2 risk categories**

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

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

Not applicable.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

Risk stratification was used rather than risk adjustment. Stratification was based on PACE site characteristics, such as census and years in operation. Because PACE participants are frail elderly in each site, they may be considered a single population, not requiring risk adjustment to account for different populations across PACE sites. Further data collection could result in additional risk stratification.
Two demographic variables—age and gender—were collected so that the potential for sociodemographic adjustment can be assessed.

- Age is defined as the participant age at the end of the reporting month. It is to be recorded in single years from 55 through 89. To comply with Health Insurance Portability and Accountability Act (HIPAA requirements), all participants aged 90 and above will be top coded at 90.
- Gender is to be classified as male or female.

We examined correlations among total fall rates and PACE site characteristics. Pearson product-moment correlation coefficient, or "*r*", was used. Pearson's *r* is a measure of the strength and direction of the linear relationship between two variables. To interpret the correlations between variables, we used the following parameters: $r = 0.80$ or higher is a very strong relationship; $r = 0.60$-0.79 is a strong relationship; $r = 0.40$-0.59 is a moderate relationship; $r = 0.20$-0.39 is a weak relationship; and $r < 0.19$ is a very weak relationship. (Evans, 1996).

Citation:
Evans, J.D. (1996). Straightforward statistics for the behavioral sciences. Pacific Grove, CA: Brooks/Cole Publishing.

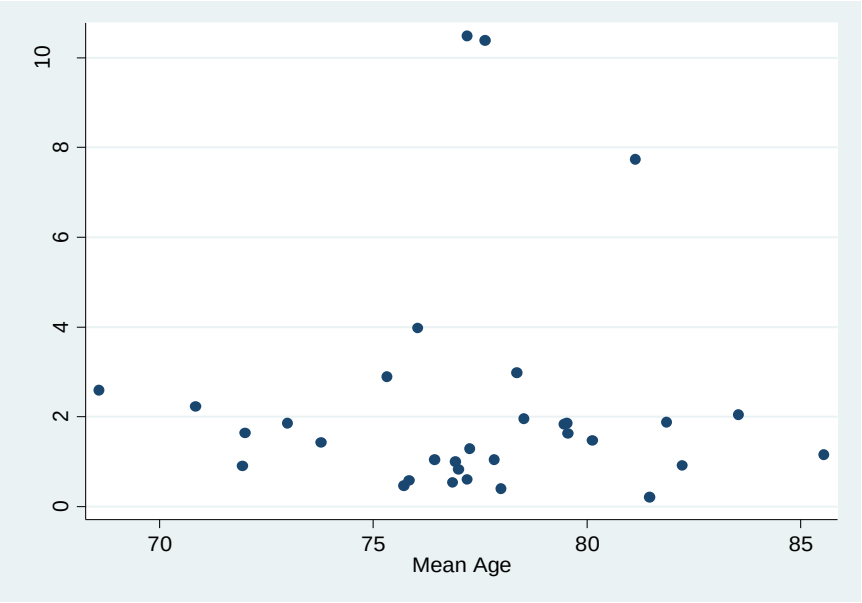
2b4.4a. What were the statistical results of the analyses used to select risk factors?

Table 1: Characteristics of Participants With Falls with Injury

Participant-Level Characteristic (*n* = 587)	Summary Statistics	
Age (years)	Mean (SD) Median (Min Max)	76.88 (10.33) 78 (56, 99)
Gender	Male Female	81 (28.32%) 205 (71.68%)
Total number of falls with injury in March 2015	Mean (SD) Median (Min Max) 1 2+	1.11 (0.37) 1 (1, 4) 258 (90.21%) 28 (0.79%)

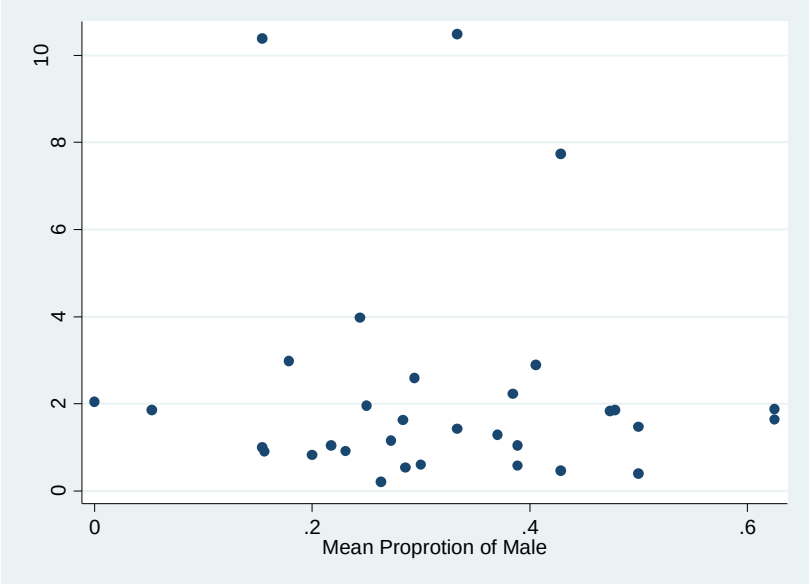
Figure 1 shows the correlations between falls with injury rates and the mean age of PACE site participants and Figure 2 shows the correlation between falls with injury rates and the mean proportion of males by PACE site. Both PACE-site participant characteristics had a very weak/negligible correlation with injury fall rates.

Figure 1: Correlation Between Falls With Injury Rates and Mean Age (r = 0.02, n = 32). Mean age was calculated by site-level mean age of participants having falls with injury.



Note: One PACE site excluded from scatterplot because of outlier data.

Figure 2: Correlation Between Falls With Injury Rates and Mean Proportion of Male (r = –0.07, n = 32). Mean proportion of male was calculated by site-level mean proportion of male having falls with injury. Negative correlation indicates that sites having more males with injurious falls were likely to have lower rates of total falls.



Note: One PACE site excluded from scatterplot because of outlier data.

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)
Both mean age and mean percent male were very weakly correlated with falls with injury rates (r = 0.02 and r = -0.07, respectively). After implementation of the measure, we will continue to collect data to determine the usefulness of risk-stratification based on age and gender.

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Not applicable.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

Age and sex were found to be only weakly correlated with falls rates (see Figures 1 and 2).

2b4.10. 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*)

Adequacy of sex and age to control for differences across PACE sites will be further evaluated after measure implementation with the larger population of PACE sites and participants.

2b4.11. 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*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE
2b5.1. 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)

Due to our small sample size, we did not conduct statistical analyses to determine differences in performance across PACE sites. However, the descriptive statistics indicate that there are differences in falls with injury rates across PACE sites (mean = 2.21, SD = 2.52, median = 1.47, range = 0.21-10.48). After implementation, we will conduct further analyses to determine significant differences in performance for falls with injury rates across PACE sites.

2b5.2. 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*)

Not applicable.

2b5.3. 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?*)

Not applicable.

2b6. 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 SDS 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 SDS 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.***

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Not Applicable.

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank*

order)

Not Applicable.

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

_ Not Applicable.

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

Because of the small amount of missing data, we did not conduct analyses of responders vs. nonresponders.

2b7.2. 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*)

In testing, 1 site reported falls, but did not report on injury falls. Because of the small amount of missing data, we did not conduct analyses of responders vs. nonresponders.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

Not applicable.

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), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

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)*

Some data elements are in defined fields in 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.**

Some PACE Organizations do not use electronic medical records. All organizations will abstract data manually for this measure from either their electronic or paper charts.

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.**

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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

After collecting data from PACE sites for feasibility and reliability testing, we conducted a post-data collection survey to ask PACE sites about data that they did not have available, data collection burden, and other issues. Overall, the data collection time was reasonable at 3-4 hours. While the sites reported a fairly high data collection burden, this was balanced by the fact that over half of the sites stated that the data were very easy to obtain. Further, all of the sites stated that fall rates are useful for quality improvement and 64% were supportive of national PACE comparison data. Thus, although there is a perceived data collection burden, this is outweighed by the usefulness of the data and comparative benchmarks. Because of the high reported ease of obtaining the data, we anticipate that the perceived data collection burden will decrease as sites become more familiar with the data collection and submission process.

- Sites said that it took between 3 and 4 hours to collect the fall rate data and another hour to submit the

data on-line.

- 73% of PACE sites reported that they considered the data collection burden to be medium or high burden.
- 69% of the sites reported that they collected falls data from electronic health records, although the large majority said they did manual extraction from electronic records.
- 54% of the sites said that it was very easy to obtain the data.
- 100% of responding sites said that the fall rates would be useful for quality improvement.
- 64% said that they strongly agreed with the statement that national comparison data would be helpful for quality improvement.

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

None.

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. **Current and Planned Use**

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

Planned	Current Use (for current use provide URL)
Public Reporting Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Not in use	

4a.1. **For each CURRENT use, checked above, provide:**

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

Not applicable.

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

This is a new measure. CMS is evaluating its use in upcoming PACE quality programs.

4a.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 is considering the use of the PACE Participant Falls With Injury Rate in accountability applications within the next two years.

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

Improvement data will be obtained once the measure has been implemented and tracked over time.

4b.2. **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.**
Improvement data will be obtained once the measure has been implemented and tracked over time.

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
No negative unintended consequences have been identified.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0202 : Falls with injury
0674 : Percent of Residents Experiencing One or More Falls with Major Injury (Long Stay)

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

5a. Harmonization

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 completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
The numerator for the falls with injury measure being developed for the PACE program is closely aligned with NQF-endorsed measures 0202. They use the same description of injury levels, however, the proposed measure uses a different denominator that reflect fall exposure in PACE programs as opposed to hospitals. NQF-endorsed measure 0266 is limited to long-stay nursing facility residents with major injuries from falls rather than any injury.

5b. Competing Measures

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

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
Not applicable.

Appendix

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

Attachment **Attachment:** AppendixA1_FallsWithInjury_Data_Collection_Sheet.docx

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** CMS
Co.2 **Point of Contact:** Stacy, Davis, stacy.davis@cms.hhs.gov, 410-786-7813-

Co.3 **Measure Developer if different from Measure Steward:** Econometrica, Inc.
Co.4 **Point of Contact:** Mark, Stewart, mstewart@econometricainc.com, 240-204-5168-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
The Falls With Injury Rate measure was developed in partnership with CMS by a team lead by Econometrica, Inc. consisting of Econometrica (prime contractor); the University Of Kansas Medical Center Research Institute (KUMCRI; subcontractor); Drs. Rosemary Kennedy and Barbara Resnick, and Ms. Heidi Bossley (consultants).

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:**
Ad.3 **Month and Year of most recent revision:**
Ad.4 **What is your frequency for review/update of this measure?**
Ad.5 **When is the next scheduled review/update for this measure?**

Ad.6 **Copyright statement:**
Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

A.1 Appendix PACE Participant Falls With Injury Rate

Participant Falls With Injury Data Abstraction Sheet

If there were no falls at your site in a month, enter 0 (zero) on the first row.

Fall No.	Month of Fall	Age (at end of month)	Gender	Injury Level
	1=January 2015 2=February 2015	Age in years if 55–89 90+=Age greater >89 99=Unknown	1=Male 2=Female 99=Unknown	1= None 2=Minor 3=Moderate 4=Major 5=Death 99=Unknowr
1				
2				
3				
4				
5				
6				

Participant Census Days

January 2015	Number of Participants in Census
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 3005

Measure Title: Initial Risk Assessment for Immobility-Related Pressure Ulcer within 24 Hours of PICU Admission

Measure Steward: Pediatric Consultants, LLC

Brief Description of Measure: This measure determines the proportion of Pediatric Intensive Care Unit (PICU) patients for whom an initial risk assessment for development of an immobility-related pressure ulcer is performed. The assessment is to be performed within the first 24 hours of admission to the PICU with the use of a standardized, validated pressure ulcer risk assessment tool designated as appropriate by the institution. The results of the assessment must be documented in the patient's chart upon completion.

Developer Rationale: A pressure ulcer is a localized injury to the skin and/or underlying tissue usually over a bony prominence that occurs as a result of pressure, or pressure in combination with shear. Pressure ulcer rates have been steadily increasing with reported rates of 4.14 pressure ulcers per 1000 pediatric discharges in 1999 and 4.33 pressure ulcers per 1000 pediatric discharges in 2002 (ref. 1). Overall, pressure ulcer incidence has increased 34.5% from 2000 to 2007 (ref. 2). Pediatric patients who experience pressure ulcers have a 6.15% mortality rate and pressure ulcers can lead to infection, pain management challenges, disfigurement, increased length of stay and readmission, altered body image, and psychological distress (ref. 2-4). Total excess cost associated with pressure ulcer patients is $1.3 billion (ref. 2). Pediatric patients with pressure ulcers experience increased hospital length of stay (mean = 8.07 days) and hospital charges (mean = $59,225) as compared to pediatric patients who do not have pressure ulcers. Excess charges occur due to pharmacy ($10,959), supplies ($4,663), laboratory ($7,276), imaging ($1,284), and other clinical activities ($11,345) (ref. 5).

Identification of patients at risk for pressure ulcer is a key step in preventing development of pressure ulcers in critically ill and injured children. Early assessment of risk has been shown to be important in the prevention of immobility-related pressure ulcer development (ref. 6-9). The Braden Q is the only validated immobility-related pressure ulcer risk assessment tool available for use with critically ill children at this time (ref. 10). Early assessment of risk using the Braden Q and/or a different validated pressure ulcer risk assessment tool can prevent the development of pressure ulcers in PICU patients, ultimately reducing morbidity and mortality rates as well as health care costs while simultaneously preventing infection and pain.

1. Sedman A, Harris JM, Schulz K, Schwalenstocker E, Remus D, Scanlon M, Bahl V. Relevance of the Agency for Healthcare Research and Quality and quality patient safety indicators for children's hospitals. Pediatrics. 2005;115:135-145.

2. Friedman B, Berdahl T, Simpson LA, McCormick MC, Owens PL, Andrews R, Romano PS. Annual report on health care for children and youth in the United States: focus on trends in hospital use and quality. Acad Pediatr. 2011;11:263-279.

3. Galvin PA, Curley MA. The Braden Q+P: a pediatric perioperative pressure ulcer risk assessment and intervention tool. AORN. 2012;96(3):261-270.

4. Baharestani MM, Ratliff CR. Pressure ulcers in neonates and children: an NPUAP white paper. Advances in Skin & Wound Care. 2007;20(4):208-220.

5. Kronman MP, Hall M, Slonim AD, Shah SS. Changes and lengths of stay attributable to adverse patient-care events using pediatric-specific quality indicators: a multicenter study of freestanding children's hospitals. Pediatrics. 2008;121(6):e1653.

6. Brandeis GH, Berlowita DR, Katz P. Are pressure ulcers preventable? A survey of experts. Advances in Skin and Wound Care. 2001;14(5):244-248.

7. Butler CT. Pediatric skin care: guidelines for assessment, prevention, and treatment. Pediatric Nursing. 2006;32(4):443-454.

8. Quigley SM, Curley MA. Skin integrity in the pediatric population: preventing and managing pressure ulcers. JSPN. 1996:1(1):7-18.

9. Sims A, McDonald R. An overview of paediatric pressure care. Journal of Tissue Viability. 2003;13:144-148.

10. Curley MA, Razmus IS, Roberts KE, Wypij D. Predicting pressure ulcer risk in pediatric patients. Nursing Research. 2003;52(1):22-31.

Numerator Statement: Number of PICU patients for whom an assessment of immobility-related pressure ulcer risk using a standardized pressure ulcer risk assessment tool was documented within 24 hours of admission.

Denominator Statement: All patients admitted to the PICU for at least 24 hours during a monthly or quarterly reporting period.

Denominator Exclusions: none

Measure Type: Process

Data Source: Electronic Clinical Data : Electronic Health Record, Other, Paper Medical Records

Level of Analysis: Facility, Integrated Delivery System

New Measure - Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *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.

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 or **Summary of prior review in [year]**

- The developers state that there are currently no clinical guidelines for pressure ulcer prevention and treatment in the pediatric population.
- Assessment tools are limited, so the Braden Q Scale was adapted from the Braden Scale of be used in this population.
- Early identification of patients at risk for pressure ulcer is a key step in preventing them in critically ill and injured children which has been shown to reduce morbidity and mortality rates as well as healthcare costs.

Guidance from the Evidence Algorithm
1-No→3-No→7-Yes→8-Yes→ MODERATE

Questions for the Committee:

- *Does the Committee agree there is no change in the evidence since the last evaluation?*

OR

If the developer provided updated evidence for this measure:

- *Questions specific to the measure information provided on evidence*
- *For process 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?*
- *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?*

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

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

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

- This measure was tested as an eMeasure at one site, Lurie Children's Hospital. Electronic output was provided for a reporting period of 01 Jan – 31 March 2015 and included 106 unique patients representing 109 events. Overall (N=106), clinical performance was high with 94% of patients meeting the measure.
- Reasons for not meeting the measure including having a pressure ulcer assessment performed outside of the 24 hour window (N=4) and not having a pressure ulcer assessment performed at all (N=3). Looking across age groups, of the children aged 0 - <6 (N=66), 92% met the measure, of the children aged 6 - <13 (N=16), 94% met the measure, of the children aged 13 - <19 (N=20), 95% met the measure, and of PICU patients 19 and older (N=4), 100% met the measure.

Disparities

- At Lurie Children's Hospital (N=106), approximately 37% (N=39) of the sample was White, 34% (N=36) was Hispanic, 16% (N=17) was Black, 12% (N=13) was Other, and less than 1% (N=1) was Unknown. The clinical performance of the eMeasure across race/ethnicity groups was as follows: 97.5% of White patients, 82% of Black patients, and 94% of Hispanic patients met the measure. Similarly, of patients who listed their race/ethnicity as "other" or "unknown", 92% and 100% met the measure, respectively. These differences were not statistically significant.

- At Lurie Children's, 61% (N=65) of patients in the sample had Private Insurance while the remaining 42% (N=41) used Medicaid. The clinical performance of the eMeasure was comparable in both groups with 95% of patients with private insurance and 90% of Medicaid patients meeting the measure criteria for having an immobility-related pressure ulcer risk assessment performed using astandardized pressure ulcer risk assessment tool within 24 hours of admission. This difference was not statistically significant.
- At Lurie Children's, 83% (N=88) of patients reported that their language preference was English, 15% (N=16) reported Spanish, and 2% (N=2) reported other languages. The clinical performance of the eMeasure across these groups was as follows: 95% of English speaking patients meeting the measure, 88% of Spanish speaking patients meeting the measure, and 100% of patients who spoke other languages meeting the measure. This difference was not statistically significant.

Questions for the Committee:

- *Specific question on information provided for gap in care.*
- *Is there a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*

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

Comments: ** This is a process measure. This involves screening in the PICU within 24 hours for the early development of pressure sores and skin breakdown using a standardized tool ("Braden Q"). The authors do provide evidence that pressure ulcer development does occur within the 24 hours of PICU stay and the incidence increases over each day thereafter. The screening for ulcers does appear to provide face validity that depending upon the score, early identification of high risk patients/skin areas should lead to earlier preventive/therapy initiation. I think this has face validity only now at this point as presented by the authors

** This is a process measure

**Yes

** This is a process measure. For a process measure, the developer is to provide a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence 4 that the measured process leads to a desired health outcome.

**There is no grading of the evidence.

**Evidence: As there are currently no clinical guidelines for pressure ulcer prevention and treatment in the pediatric population and assessment tools are limited, the developers reported that the Braden Q Scale was adapted from the Braden Scale to be used in this population. The predictive validity of the Braden Q Scale was established in an acutely ill pediatric population and the critical cutoff points for classifying patient risk were determined as well as the best time to assess patient risk.

1b. Performance Gap

Comments: ** The authors do not present strong gap data as to whether there is either a current practice variation in the documentation (usually part of basic nursing care) of skin integrity or their tool identifies patients more than without or whether their tool adds a superiority.

** While I accept the idea that pressure sores are a significant risk for pediatric patients, I am having problems with the data used to support this finding. Specifically, I am uneasy about total reliance on one and only one measurement protocol (Braden) and need reassurance/evidence that this discussion is free of market interests for the makers of the Braden protocol. I would be far more comfortable with this discussion if several protocols were available.

** The references are very dated: Pressure ulcer rates have been steadily increasing with reported rates of 4.14 pressure ulcers per 1000 pediatric discharges in 1999 and 4.33 pressure ulcers per 1000 pediatric discharges in 2002.2 Overall, pressure ulcer incidence has increased 34.5% from 2000 to 2007.3

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s):

Specifications:

- This measure assesses the proportion of PICU patients for whom an initial risk assessment for development of an immobility-related pressure ulcer has been performed within 24 hours of admission.
- The measure is specified at the hospital facility or integrated delivery system level of analysis, and is meant to be reported on a monthly or quarterly basis.
- The denominator includes all patients admitted to the PICU for at least 24 hours during the reporting period.
- The numerator includes patients from the denominator population who have been assessed for risk of pressure ulcers using a standardized, validated tool.
    - The measure defines a standardized, validated pressure ulcer risk assessment tool as "a validated assessment tool that is applied in a standardized fashion to each patient admitted to the PICU for at least 24 hours."
    - The developer notes that, currently, the Braden Q is the only validated immobility-related pressure ulcer risk assessment tool available for critically ill and injured children; however, the measure allows for the use of other validated risk assessment tools, if available.
- The measure is specified as an eMeasure; a technical review of the eMeasure specifications is included below.

Questions for the Committee :

- *Specific questions on the specifications, codes, definitions, etc.*
- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

eMeasure Technical Advisor(s) review :

Submitted measure is an HQMF compliant eMeasure	The submitted eMeasure specifications follow the industry accepted format for eMeasure (HL7 Health Quality Measures Format (HQMF)). **HQMF specifications** ☒ **Yes** ☐ **No**
Documentation of HQMF or QDM limitations	N/A – All components in the measure logic of the submitted eMeasure are represented using the HQMF and QDM; **OR** **All components of the HQMF were defined and clearly specified. There was a defined element that mapped to the HQMF tag that was inclusive of both the numerator and denominator. The developer chose to create their own QDM elements under User Defined QDM Value Set (1.1.1.1) and set them under the Encounter and Risk Assessment headings. This is perfectly appropriate as the QDM elements have to be aligned to the measure and the appropriate components, which they are in this case.** Submitted eMeasure contains components that cannot be represented due to limitations of HQMF or QDM and the submission explains the work around for these limitations; **OR** Submitted eMeasure contains components that cannot be represented due to limitations HQMF or QDM and the submission does NOT explain the work around for these limitations

Value Sets	The submitted eMeasure specifications uses existing value sets when possible and uses new value sets that have been vetted through the VSAC **OR** **The value sets included in this measure are published within the VSAC.** Some value sets used in the submitted eMeasure are not present in the NLM Value Set Authority Center but the measure developer has provided justification for using such value sets
Measure logic is unambiguous	Submission includes test results from a simulated data set demonstrating the measure logic can be interpreted precisely and unambiguously; **OR** **The developer used data sets from two EHR systems: Epic and Cerner, and were able to show results that mapped to the logic from both systems. However, Occurrence of PICU admission is not a structured data field that is mapped to a national vocabulary, which may cause some issue in implementation.** Submission does not include test results from a simulated data set demonstrating the measure logic can be interpreted precisely and unambiguously; **OR** Submission includes test results from a simulated data set demonstrating the measure logic cannot be interpreted precisely and unambiguously.
Feasibility Testing	The submission contains a feasibility assessment that addresses data element feasibility and follow-up with measure developer indicates that the measure logic is feasible based on assessment by EHR vendors**; OR** **The feasibility scorecard is completed and the scores represented high feasibility with the note that Occurrence of PICU Admission is not a nationally coded element.** The feasibility analysis submitted by the measure developer meets the requirements to be considered for eMeasure Trial Approval.

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☒ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☐ **Yes** ☐ **No**

Method(s) of reliability testing

- While eMeasure feasibility testing was conducted in four Chicago-area hospitals, reliability testing was only conducted in one of those hospitals because of implementation issues at the other three hospitals.
- To demonstrate reliability, the developer performed data element testing at one hospital site with 288 pediatric beds (including 40 PICU beds) and approximately 11,291 pediatric admissions annually.
- Patients were included in the testing if they were admitted to the PICU during 01 Jan – 31 March 2015 at Lurie Children's Hospital; the analysis included 106 unique patients, representing 109 events.
- The testing involved implementation of the eMeasure to compute scores automatically, and manual chart review of the same patients by a trained chart abstracter; inter-rater reliability was then assessed.

Results of reliability testing [Results of reliability testing]

- The developer reports that inter-rater reliability was 100% for all critical data elements, and 100% for overall clinical performance of the measure.
- Because agreement was 100%, a Kappa score could not be computed.
- The developer interprets the results as indicating that the measure has good reliability when compared to the gold standard of chart reviews.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

[Box 1] Specifications precise and unambiguous → [Box 2] Empirical testing NOT conducted on the measure as specified → [Box 3] Empirical validity testing of patient-level data conducted → [Box 11 of **Validity Algorithm**] Testing method described and appropriate → [Box 12] Moderate certainty or confidence that the data used in the measure are valid (some concern about testing only being conducted in a single site)

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

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

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.

SUMMARY OF TESTING

Validity testing level ☐ **Measure score** ☒ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

- ☐ **Face validity only**
- ☐ **Empirical validity testing of the measure score**

Validity testing method:

- See reliability section above (data element validity testing)

Validity testing results:

- See reliability section above (data element validity testing)

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*
- *Other specific question of the validity testing?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

N/A

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

N/A

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- To assess whether measure scores can detect meaningful differences in performance, the developer used Chi-square statistics to determine if there were statistically significant differences between age groups, race/ethnicities, health insurance plans (private vs. Medicaid), and preferred language.
- The developer reported that there were no statistically significant and clinically/practically meaningful differences in performance.
- Interpreting this information, the developer states that, "as there were no statistically significant differences between measured entities, we can conclude that measure performance was similar across patient factors and hospitals."
- The developer did not analyze the measure's ability to detect meaningful differences in performance across facilities.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

N/A

2b7. Missing Data

- With regard to missing data, the developer reports that, in order to meet the denominator criteria for the measure, all components of the denominator must be present in the patient chart.
- If numerator data is missing, it is assumed that the care element was not provided and the patient chart does not meet numerator criteria.
- In data used for measure testing, 109 events were identified and of those events, all met the denominator criteria. Only three events (2.75%) were missing information on pressure ulcer risk assessment and they were assumed to have not been performed.

Guidance from the Validity Algorithm

[Box 1] Specifications consistent with evidence → [Box 2] Potential threats to validity addressed, but meaningful differences information is questionable → [Box 3] Empirical validity testing conducted using the measure as specified → [Box 6] Validity testing NOT conducted with computed measure scores → [Box 10] Validity testing conducted with patient-level data elements [Box 11] Testing method described and appropriate → [Box 12] Moderate certainty or confidence that the data used in the measure are valid (some concern about testing only being conducted in a single site)

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

Measure received a preliminary rating of "insufficient" because data were not provided to demonstrate that the measure identifies meaningful differences in performance.

Committee pre-evaluation comments

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

2a1. & 2b1. Specifications

Comments: ** Screening can improve care but the tool recommended in this measure involves 28 data choices in 7 categories which could result in scores ranging from 7-28. The specifications as to what entails a complete screen arwe not presented

** The measure is specified at the hospital level and assesses the proportion of PICU patients for whom an initial risk assessment for development of a immobility related pressure ulcer is done within 24 hours of admission. Specifications are clear and consistent with evidence. HQMF specifications are followed
** Reliability and specifications seem clear.
** Numerator Statement: Number of PICU patients for whom an assessment of immobility-related pressure ulcer risk using a standardized pressure ulcer risk assessment tool was documented within 24 hours of admission.
Denominator Statement: All patients admitted to the PICU for at least 24 hours during a monthly or quarterly reporting period.

Denominator Exclusions: none (PQ – question this)
If the Assessment has to be documented within 24 hours, would not an exclusion be those admitted for less that 24 hours? Why is this not expected within the first 8 or 12 hour shift of admission?

Is this not a standard of care for pediatrics? This is nursing practice.

Analysis Level is facility- Shouldn't this be at the Unit Level

Is the template in all EMRs?

2a2. Reliability Testing
Comments: ** Data element reliability testing was done in one hospital with 288 pediatric beds, including 40 picu beds and 11.291 ped admissions annually -- over a three month period. Implementation of the eMeasure was tested by having scores computer automatically and having manual chart review of the same patients done. Inter-rater reliability was 100%.
** Method(s) of reliability testing
**While eMeasure feasibility testing was conducted in four Chicago-area hospitals, reliability testing was only conducted in one of those hospitals because of implementation issues at the other three hospitals.
**To demonstrate reliability, the developer performed data element testing at one hospital site with 288 pediatric beds (including 40 PICU beds) and approximately 11,291 pediatric admissions annually.
**Patients were included in the testing if they were admitted to the PICU during 01 Jan – 31 March 2015 at Lurie Children's Hospital; the analysis included 106 unique patients, representing 109 events.
** The developer reports that inter-rater reliability was 100% for all critical data elements, and 100% for overall clinical

2b2. Validity Testing
Comments: ** I do not think the authors show that validity. Does one have to do the complete score or part of it? If this measure is aver and above the normal care - no data is present to show it adds marginal benefit
** Validity testing was done on the data element against a gold standard -- see reliability section.
**
Validity The Braden Q is the only validated immobility-related pressure ulcer risk assessment tool available for use with critically ill children at this time (ref. 10). Early assessment of risk using the Braden Q and/or a different validated pressure ulcer risk assessment tool can prevent the development of pressure ulcers in PICU patients, ultimately reducing morbidity and mortality rates as well as health care costs while simultaneously preventing infection and pain.

The references are very dated: 2 – 2003, and 1 – 2011

Curley MA, Razmus IS, Roberts KE, Wypij D. Predicting pressure ulcer risk in pediatric patients: the Braden Q Scale. Nurs Res. 2003;52(1):22-33.

The performance of the Braden Q in pediatric patients was determined to be similar to that consistently reported for the Braden Scale in adult patients. A cutoff score of 16 provided a high sensitivity and adequate specificity (0.88 and 0.58, respectively). This produced a Likelihood Ratio of 2.11. Most pressure ulcers developed soon after PICU admission and it was established that the best time to assess a pediatric patient for pressure ulcer risk is within 24 hours of PICU admission.

2b3. Exclusions Analysis
2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures
2b5. Identification of Statistically Significant & Meaningful Differences In Performance
2b6. Comparability of Performance Scores When More Than One Set of Specifications
2b7. Missing Data Analysis and Minimizing Bias

Comments: ** No patients are excluded - this is appropriate. The validity threat is in the details of the scoring and any inter-rater reliability
** Chi-square statistics were used to determine if there were statistically different differences between age groups, race/ethnicity, health insurance plans and preferred language. None were identified. The developer did not analyze the measure's ability to detect meaningful differences in performance across facilities.
With regard to missing data, to be counted in the denominator, all components of the denominator must be present. If numerator data is missing, it is assumed the care element was not provided. Only 2.75% were missing information in measure testing.
**
Threats to validity
**With regard to missing data, the developer reports that, in order to meet the denominator criteria for the measure, all components of the denominator must be present in the patient chart.
**If numerator data is missing, it is assumed that the care element was not provided and the patient chart does not meet numerator criteria.
**In data used for measure testing, 109 events were identified and of those events, all met the denominator criteria. Only three events (2.75%) were missing information on pressure ulcer risk assessment and they were assumed to have not been performed.

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.

- This measure is generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score)
- Braden Q is listed as the only validated immobility-related pressure ulcer risk assessment tool available for use
- eMeasure Feasibility Scorecard included
  - The developer included feasibility scorecard results for EHR records from Epic and Cerner.
  - However, the developer did not provide OID's listed for any of the four data elements listed.
- The developer stated that feasibility testing was conducted on four sites and the measure was able to be implemented in three sites, however the developer only provided feasibility scorecards for two hospitals.
- The MIF reflects that ALL data elements are in defined fields in electronic health records (EHRs)
- The scorecard indicated that at two sites (the Hospital Network) data elements are routinely collected as part of routine care and no additional data entry was required. At one hospital, while the pressure ulcer risk assessment tool was built within their HER system, clinicians did not routine use it, preferring to perform the risk assessment on paper based forms and scanning them into the EHR. The developer has hope that within the next year, hospitals will use the electronic risk assessment tool, already in existence in the EHR, to perform this assessment.
- No fees, licensing, or other requirements to use any aspect of the measure as specified were identified.

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?*
- *If an eMeasure, does the eMeasure Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites?*
- *Is it important to know how feasible the Braden Q scale was, vs. 'other sources'?*

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

Committee pre-evaluation comments
Criteria 3: Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: ** The measure itself is good and well intentioned. The tool itself is more detailed than the authors lead one to believe (even if scored a simple yes/no as to done - this score is actually 28 elements in 7 categories and at least one article I read in my own search highlighted the importance of education. Within an EMR this may be more feasible but in an PICU setting on those who need most, the detail of the Braden Q may be quite cumbersome

** Most data is available in EHRs. Braden Q is the only validated tool available. Feasibility testing was done at 4 hospitals; the measure was able to be implemented in 3, but scorecards were only provided for 2 hospitals.

** If documentation issues can be clarified, the feasibility of a measure for occurrence of pressure ulcers should be feasible. It would be part of the chart and also part of billing for treatment.

** Feasibility:

- This measure is generated or collected by and used by healthcare personnel during the provision of care
- Braden Q is listed as the only validated immobility-related pressure ulcer risk assessment tool available for use
- eMeasure Feasibility Scorecard included
- The developer included feasibility scorecard results for EHR records from Epic and Cerner.
- However, the developer did not provide OID's listed for any of the four data elements listed.
- The developer stated that feasibility testing was conducted on four sites and the measure was able to be implemented in three sites, however the developer only provided feasibility scorecards for two hospitals.
- The MIF reflects that ALL data elements are in defined fields in electronic health records (EHRs) – is this all EHRs?
- The scorecard indicated that at two sites (the Hospital Network) data elements are routinely collected as part of routine care and no additional data entry was required.
- At one hospital, while the pressure ulcer risk assessment tool was built within their HER system, clinicians did not routine use it, preferring to perform the risk assessment on paper based forms and scanning them into the EHR.
- The developer has hope that within the next year, hospitals will use the electronic risk assessment tool, already in existence in the EHR, to perform this assessment.

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

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

Current uses of the measure [from OPUS]
Publicly reported? ☐ **Yes** ☒ **No**

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

- This measure is not currently in use, however, the developer provided a credible plan for implementation with the expected timeframes.
- This measure is being submitted for endorsement for use in public and private health plans, Medicaid, and CHIPRA to assess the quality of care related to the prevention of pressure ulcers for children in the PICU for public reporting and quality improvement.
- The developer sees this measure becoming a part of an American Board of Pediatrics (ABP) Maintenance of Certification (MOC) Performance Improvement Module (PIM).
- The developer also foresees this measure being tested as a discrete module in the Virtual Pediatric System (VPS) pending receipt of funding from AHRQ.

There were no Unexpected findings (positive or negative) during implementation

There were no unintended negative consequences to individuals or populations identified during testing.

***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

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: ** It is not being publically reported at this time

** This measure is not publicly reported. and is not used in an accountability program. The developer provided a plan for implementation with expected timeframes in private health plans, Medicaid, CHIPRA.

** If this measure were publicly reported, it would be useful for families to know. If the measure is found to be useful, I would push for public disclosure.

** Usability: No currently reported, not used in an accountability program, but planned use in an accountability program; This measure is being submitted for endorsement for use in public and private health plans, Medicaid, and CHIPRA to assess the quality of care related to the prevention of pressure ulcers for children in the PICU for public reporting and quality improvement.

Criterion 5: Related and Competing Measures

Related or competing measures

- 0337 : Pressure Ulcer Rate (PDI 2)
- 0539 : Pressure Ulcer Prevention Implemented during Short Term Episodes of Care

Harmonization

- NQF measure #0539, Pressure Ulcer Prevention and Care, is a pressure ulcer prevention measure targeted towards the adult population in a home health setting. While this measure appears to be somewhat comparable to the PICU this measure is designed for critically ill and injured children in the PICU, an entirely different patient population and medical care setting.
- NQF measure #0337, Pressure Ulcer Rate (PDI2), is a measure that is targeted at the same age group as this proposed measure, the current endorsed measure assesses the percentage of patients who have a Stage III or IV pressure ulcer. This measure requires the use of a validated tool to assess immobility pressure ulcer risk in order to prevent the occurrence of developing a pressure ulcer at all. This measure is applied only to the care of critically ill and injured children in the PICU, a more circumscribed, but more at risk population.

Pre-meeting public and member comments

- N/A

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

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

Measure Title: Initial Risk Assessment for Immobility-related Pressure Ulcer within 24 hours of PICU Admission

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

Date of Submission: 5/13/2016

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

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

Outcome

☐ Health outcome: Click here to name the health outcome

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

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

☒ Process: Immobility-related pressure ulcer risk assessment

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

N/A

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

N/A

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

Pressure ulcers develop when soft tissue (muscle, fat, fibrous tissue, blood vessels, or other supporting tissue of the body) is compressed between a bony prominence and an external surface for a prolonged period of time. The ulcer forms when arterioles and capillaries collapse under this external pressure, resulting in a limited oxygen supply and a decrease in the transportation of vital nutrients to the cells. This results in tissue hypoxia, causing cellular death, injury to the surrounding area, and ultimately a pressure ulcer.[1]

A pressure ulcer is a localized injury to the skin and/or underlying tissue usually over a bony prominence that occurs as a result of pressure, or pressure in combination with shear.[2] Pressure ulcer rates have been steadily increasing with reported rates of 4.14 pressure ulcers per 1000 pediatric discharges in 1999 and 4.33 pressure ulcers per 1000 pediatric discharges in 2002.[2] Overall, pressure ulcer incidence has increased 34.5% from 2000 to 2007.[3] Pediatric patients who experience pressure ulcers have a 6.15% mortality rate and pressure ulcers can lead to infection, pain management challenges, disfigurement, increased length of stay and readmission, altered body image, and psychological distress.[3-5] Total excess cost associated with pressure ulcer patients is $1.3 billion.[3] Pediatric patients with pressure ulcers experience increased hospital length of stay (mean = 8.07 days) and hospital charges (mean = $59,225) as compared to pediatric patients who do not have pressure ulcers. Excess charges occur due to pharmacy ($10,959), supplies ($4,663), laboratory ($7,276), imaging ($1,284), and other clinical activities ($11,345).[6]

Early intervention can be an effective prevention measure against pressure ulcer development. Pressure ulcer prevention begins with accurate assessment to identify at-risk patients. The Braden-Q is the only validated immobility-related pressure ulcer risk assessment tool available for critically ill and injured children. The Braden Q consists of seven subscales: mobility, activity, sensory perception, moisture, friction/shear, nutrition, and tissue perfusion/oxygenation. When an assessment identifies pressure ulcer risk as high, interventions, such as patient repositioning, should be implemented to reduce the risk of pressure ulcer development.[1]

Identifying patients at-risk for pressure ulcer and then intervening accordingly can reduce the incidence of immobility-related pressure ulcer development which ultimately reduces infection, pain, disfigurement, length of stay, readmission, psychological distress, and mortality in PICU patients.[7,8]

1. Butler CT. Pediatric skin care: guidelines for assessment, prevention, and treatment. Pediatric Nursing. 2006;32(5):443-454.
2. Sedman A, Harris JM, Schulz K, Schwalenstocker E, Remus D, Scanlon M, Bahl V. Relevance of the Agency for Healthcare Research and Quality and quality patient safety indicators for children's hospitals. Pediatrics. 2005;115:135-145.
3. Friedman B, Berdahl T, Simpson LA, McCormick MC, Owens PL, Andrews R, Romano PS. Annual report on health care for children and youth in the United States: focus on trends in hospital use and quality. Acad Pediatr. 2011;11:263-279.
4. Galvin PA, Curley MA. The Braden Q+P: a pediatric perioperative pressure ulcer risk assessment and intervention tool. AORN. 2012;96(3):261-270.
5. Baharestani MM, Ratliff CR. Pressure ulcers in neonates and children: an NPUAP white paper. Advances in Skin & Wound Care. 2007;20(4):208-220.
6. Kronman MP, Hall M, Slonim AD, Shah SS. Changes and lengths of stay attributable to adverse patient-care events using pediatric-specific quality indicators: a multicenter study of freestanding children's hospitals. Pediatrics. 2008;121(6):e1653.
7. Galvin PA, Curley MA. The Braden Q+P: a pediatric perioperative pressure ulcer risk assessment and intervention tool. AORN. 2012;96(3):261-270.
8. Baharestani MM, Ratliff CR. Pressure ulcers in neonates and children: an NPUAP white paper. Advances in Skin & Wound Care. 2007;20(4):208-220.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☒ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: Click here to enter date range

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for each new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

1a.8 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.8.1 What process was used to identify the evidence?

As there are currently no clinical guidelines for pressure ulcer prevention and treatment in the pediatric population and assessment tools are limited, the Braden Q Scale was adapted from the Braden Scale to be used in this population. The predictive validity of the Braden Q Scale was established in an acutely ill pediatric population and the critical cutoff points for classifying patient risk were determined as well as the best time to assess patient risk.

1a.8.2. Provide the citation and summary for each piece of evidence.

Curley MA, Razmus IS, Roberts KE, Wypij D. Predicting pressure ulcer risk in pediatric patients: the Braden Q Scale. Nurs Res. 2003;52(1):22-33.

The performance of the Braden Q in pediatric patients was determined to be similar to that consistently reported for the Braden Scale in adult patients. A cutoff score of 16 provided a high sensitivity and adequate specificity (0.88 and 0.58, respectively). This produced a Likelihood Ratio of 2.11. Most pressure ulcers developed soon after PICU admission and it was established that the best time to assess a pediatric patient for pressure ulcer risk is within 24 hours of PICU admission.

Noonan C, Quigley S, Curley M. Using the Braden Q Scale to predict pressure ulcer risk in pediatric patients. J Pediatr Nurs. 2011; 26:566-575.

In order to prevent pressure ulcer development, at a minimum, a complete skin assessment that includes a Braden Q Scale score should be completed within 24 hours of admission. Experts recommend that the Braden Q Scale be repeated daily on all patients who score 16 or less, are on bed rest or chairfast, or who have a change in clinical condition. The patient considered “at risk” for pressure ulcer development should have risk reduction

interventions put into place to minimize risk. Once interventions are in place, the patient's risk can be reassessed and scored periodically with interventions appropriately removed as the score improves.

Pressure ulcers in pediatric patients have been reported to occur by a patient's second hospital day. Recent changes in federal regulations have highlighted the importance of assessing patients' risk for pressure ulcers as soon possible upon hospital admission. Since October 2008, the Centers for Medicare & Medicaid Services "present-on-admission" pressure ulcer regulations consider any pressure ulcer that is not documented within 24 hours of admission to be considered hospital acquired.

Curley MAQ, Quigley SM, Lin M. Pressure ulcers in pediatric intensive care: incidence and associated factors. Pediatr Crit Care Med. 2003; 4(3):284-290.

Most pressure ulcers (57%) were present at the first observation period on PICU day 2 (median, 1; IQR, 1-2); all but one pressure ulcer developed before the fourth observation period (day 8).

Measure Information

This document contains the information submitted by measure developers/stewards, but is organized according to NQF's measure evaluation criteria and process. The item numbers refer to those in the submission form but may be in a slightly different order here. In general, the item numbers also reference the related criteria (e.g., item 1b.1 relates to subcriterion 1b).

Brief Measure Information

NQF #: 3005
Corresponding Measures:
De.2. **Measure Title:** Initial Risk Assessment for Immobility-Related Pressure Ulcer within 24 Hours of PICU Admission
Co.1.1. **Measure Steward:** Pediatric Consultants, LLC
De.3. **Brief Description of Measure:** This measure determines the proportion of Pediatric Intensive Care Unit (PICU) patients for whom an initial risk assessment for development of an immobility-related pressure ulcer is performed. The assessment is to be performed within the first 24 hours of admission to the PICU with the use of a standardized, validated pressure ulcer risk assessment tool designated as appropriate by the institution. The results of the assessment must be documented in the patient's chart upon completion.
1b.1. **Developer Rationale:** A pressure ulcer is a localized injury to the skin and/or underlying tissue usually over a bony prominence that occurs as a result of pressure, or pressure in combination with shear. Pressure ulcer rates have been steadily increasing with reported rates of 4.14 pressure ulcers per 1000 pediatric discharges in 1999 and 4.33 pressure ulcers per 1000 pediatric discharges in

2002 (ref. 1). Overall, pressure ulcer incidence has increased 34.5% from 2000 to 2007 (ref. 2). Pediatric patients who experience pressure ulcers have a 6.15% mortality rate and pressure ulcers can lead to infection, pain management challenges, disfigurement, increased length of stay and readmission, altered body image, and psychological distress (ref. 2-4). Total excess cost associated with pressure ulcer patients is $1.3 billion (ref. 2). Pediatric patients with pressure ulcers experience increased hospital length of stay (mean = 8.07 days) and hospital charges (mean = $59,225) as compared to pediatric patients who do not have pressure ulcers. Excess charges occur due to pharmacy ($10,959), supplies ($4,663), laboratory ($7,276), imaging ($1,284), and other clinical activities ($11,345) (ref. 5).

Identification of patients at risk for pressure ulcer is a key step in preventing development of pressure ulcers in critically ill and injured children. Early assessment of risk has been shown to be important in the prevention of immobility-related pressure ulcer development (ref. 6-9). The Braden Q is the only validated immobility-related pressure ulcer risk assessment tool available for use with critically ill children at this time (ref. 10). Early assessment of risk using the Braden Q and/or a different validated pressure ulcer risk assessment tool can prevent the development of pressure ulcers in PICU patients, ultimately reducing morbidity and mortality rates as well as health care costs while simultaneously preventing infection and pain.

1. Sedman A, Harris JM, Schulz K, Schwalenstocker E, Remus D, Scanlon M, Bahl V. Relevance of the Agency for Healthcare Research and Quality and quality patient safety indicators for children's hospitals. Pediatrics. 2005;115:135-145.

2. Friedman B, Berdahl T, Simpson LA, McCormick MC, Owens PL, Andrews R, Romano PS. Annual report on health care for children and youth in the United States: focus on trends in hospital use and quality. Acad Pediatr. 2011;11:263-279.

3. Galvin PA, Curley MA. The Braden Q+P: a pediatric perioperative pressure ulcer risk assessment and intervention tool. AORN. 2012;96(3):261-270.

4. Baharestani MM, Ratliff CR. Pressure ulcers in neonates and children: an NPUAP white paper. Advances in Skin & Wound Care. 2007;20(4):208-220.

5. Kronman MP, Hall M, Slonim AD, Shah SS. Changes and lengths of stay attributable to adverse patient-care events using pediatric-specific quality indicators: a multicenter study of freestanding children's hospitals. Pediatrics. 2008;121(6):e1653.

6. Brandeis GH, Berlowita DR, Katz P. Are pressure ulcers preventable? A survey of experts. Advances in Skin and Wound Care. 2001;14(5):244-248.

7. Butler CT. Pediatric skin care: guidelines for assessment, prevention, and treatment. Pediatric Nursing. 2006;32(4):443-454.

8. Quigley SM, Curley MA. Skin integrity in the pediatric population: preventing and managing pressure ulcers. JSPN. 1996:1(1):7-18.

9. Sims A, McDonald R. An overview of paediatric pressure care. Journal of Tissue Viability. 2003;13:144-148.

10. Curley MA, Razmus IS, Roberts KE, Wypij D. Predicting pressure ulcer risk in pediatric patients. Nursing Research. 2003;52(1):22-31.

S.4. **Numerator Statement:** Number of PICU patients for whom an assessment of immobility-related pressure ulcer risk using a standardized pressure ulcer risk assessment tool was documented within 24 hours of admission.
S.7. **Denominator Statement:** All patients admitted to the PICU for at least 24 hours during a monthly or quarterly reporting period.
S.10. **Denominator Exclusions:** none

De.1. **Measure Type:** Process
S.23. **Data Source:** Electronic Clinical Data : Electronic Health Record, Other, Paper Medical Records
S.26. **Level of Analysis:** Facility, Integrated Delivery System

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, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
Evidence_Attachment_-_Pressure_Ulcer_5.12.16.docx

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., the benefits or improvements in quality envisioned by use of this measure)*
A pressure ulcer is a localized injury to the skin and/or underlying tissue usually over a bony prominence that occurs as a result of pressure, or pressure in combination with shear. Pressure ulcer rates have been steadily increasing with reported rates of 4.14 pressure ulcers per 1000 pediatric discharges in 1999 and 4.33 pressure ulcers per 1000 pediatric discharges in 2002 (ref. 1). Overall, pressure ulcer incidence has increased 34.5% from 2000 to 2007 (ref. 2). Pediatric patients who experience pressure ulcers have a 6.15% mortality rate and pressure ulcers can lead to infection, pain management challenges, disfigurement, increased length of stay and readmission, altered body image, and psychological distress (ref. 2-4). Total excess cost associated with pressure ulcer patients is $1.3 billion (ref. 2). Pediatric patients with pressure ulcers experience increased hospital length of stay (mean = 8.07 days) and hospital charges (mean = $59,225) as compared to pediatric patients who do not have pressure ulcers. Excess charges occur due to pharmacy ($10,959), supplies ($4,663), laboratory ($7,276), imaging ($1,284), and other clinical activities ($11,345) (ref. 5).

Identification of patients at risk for pressure ulcer is a key step in preventing development of pressure ulcers in critically ill and injured children. Early assessment of risk has been shown to be important in the prevention of immobility-related pressure ulcer development (ref. 6-9). The Braden Q is the only validated immobility-related pressure ulcer risk assessment tool available for use with critically ill children at this time (ref. 10). Early assessment of risk using the Braden Q and/or a different validated pressure ulcer risk assessment tool can prevent the development of pressure ulcers in PICU patients, ultimately reducing morbidity and mortality rates as well as health care costs while simultaneously preventing infection and pain.

1. Sedman A, Harris JM, Schulz K, Schwalenstocker E, Remus D, Scanlon M, Bahl V. Relevance of the Agency for Healthcare Research and Quality and quality patient safety indicators for children's hospitals. Pediatrics. 2005;115:135-145.

2. Friedman B, Berdahl T, Simpson LA, McCormick MC, Owens PL, Andrews R, Romano PS. Annual report on health care for children and youth in the United States: focus on trends in hospital use and quality. Acad Pediatr. 2011;11:263-279.

3. Galvin PA, Curley MA. The Braden Q+P: a pediatric perioperative pressure ulcer risk assessment and intervention tool. AORN. 2012;96(3):261-270.

4. Baharestani MM, Ratliff CR. Pressure ulcers in neonates and children: an NPUAP white paper. Advances in Skin & Wound Care. 2007;20(4):208-220.

5. Kronman MP, Hall M, Slonim AD, Shah SS. Changes and lengths of stay attributable to adverse patient-care events using pediatric-specific quality indicators: a multicenter study of freestanding children's hospitals. Pediatrics. 2008;121(6):e1653.

6. Brandeis GH, Berlowita DR, Katz P. Are pressure ulcers preventable? A survey of experts. Advances in Skin and Wound Care. 2001;14(5):244-248.

7. Butler CT. Pediatric skin care: guidelines for assessment, prevention, and treatment. Pediatric Nursing. 2006;32(4):443-454.

8. Quigley SM, Curley MA. Skin integrity in the pediatric population: preventing and managing pressure ulcers. JSPN. 1996:1(1):7-18.

9. Sims A, McDonald R. An overview of paediatric pressure care. Journal of Tissue Viability. 2003;13:144-148.

10. Curley MA, Razmus IS, Roberts KE, Wypij D. Predicting pressure ulcer risk in pediatric patients. Nursing Research. 2003;52(1):22-31.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
This measure was tested as an eMeasure at one site, Lurie Children's Hospital. Electronic output was provided for a reporting period of 01 Jan – 31 March 2015 and included 106 unique patients representing 109 events. Overall (N=106), clinical performance was high with 94% of patients meeting the measure. Reasons for not meeting the measure including having a pressure ulcer assessment performed outside of the 24 hour window (N=4) and not having a pressure ulcer assessment performed at all (N=3). Looking across age groups, of the children aged 0 - <6 (N=66), 92% met the measure, of the children aged 6 - <13 (N=16), 94% met the measure, of the children aged 13 - <19 (N=20), 95% met the measure, and of PICU patients 19 and older (N=4), 100% met the measure.

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*
At Lurie Children's Hospital (N=106), approximately 37% (N=39) of the sample was White, 34% (N=36) was Hispanic, 16% (N=17) was Black, 12% (N=13) was Other, and less than 1% (N=1) was Unknown. The clinical performance of the eMeasure across race/ethnicity groups was as follows: 97.5% of White patients, 82% of Black patients, and 94% of Hispanic patients met the measure. Similarly, of patients who listed their race/ethnicity as "other" or "unknown", 92% and 100% met the measure, respectively. These differences were not statistically significant.

At Lurie Children's, 61% (N=65) of patients in the sample had Private Insurance while the remaining 42% (N=41) used Medicaid. The clinical performance of the eMeasure was comparable in both groups with 95% of patients with private insurance and 90% of Medicaid patients meeting the measure criteria for having an immobility-related pressure ulcer risk assessment performed using a standardized pressure ulcer risk assessment tool within 24 hours of admission. This difference was not statistically significant.

At Lurie Children's, 83% (N=88) of patients reported that their language preference was English, 15% (N=16) reported Spanish, and 2% (N=2) reported other languages. The clinical performance of the eMeasure across these groups was as follows: 95% of English speaking patients meeting the measure, 88% of Spanish speaking patients meeting the measure, and 100% of patients who spoke other languages meeting the measure. This difference was not statistically significant.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
n/a

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
A leading cause of morbidity/mortality, High resource use, Patient/societal consequences of poor quality
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
Pressure ulcers pose significant health problems in the PICU (ref. 1). Pressure ulcer incidence in critically ill infants and children has been reported to be as high as 18% to 27% (ref.2). Among PICU patients, approximately 8.5% experience skin breakdown, 6.2% experience redness, and 3.2% experience both redness and skin breakdown. Patients who have skin breakdown and redness are more likely to be younger, experience longer hospital stays, and are more likely to have respiratory illnesses and require mechanical ventilator support. They also have a higher mortality rate (ref.2). Provided the high incidence rate of pressure ulcer and the fact that PICU patients who experience pressure ulcers have increased mortality rates and decreased quality of life, pressure ulcer prevention is a high priority aspect of healthcare for pediatric patients. Use of a validated immobility related pressure ulcer risk assessment tool can identify patients at risk in order to prevent the development of a pressure ulcer, thereby reducing the risks of morbidity and mortality in this vulnerable population from the development of a pressure ulcer.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
1. Schindler CA, Mikhailov TA, Fischer K, Lukasiewicz G, Kuhn E, Duncan L. Skin integrity in critically ill and injured children. Am J Crit Care. 2006;16(6):568-574.

2. Schindler CA, Mikhailov TA, Kuhn EM, Christopher J, Conway P, Ridling D, Simpson VS. Protecting fragile skin: nursing interventions to decrease development of pressure ulcers in pediatric intensive care. Am J Crit Care. 2011;20(1):26-34.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
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 subcriteria 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):*
Prevention, Prevention : Screening, Pulmonary/Critical Care, Pulmonary/Critical Care : Critical Care

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Prevention, Prevention : Screening, Safety, Safety : Complications

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.)*
http://www.ahrq.gov/sites/default/files/wysiwyg/policymakers/chipra/factsheets/chipra-16-p002-1-ef.pdf

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 an eMeasure **Attachment:** S.2a._Measure_Specs_-_Pressure_Ulcer.pdf,PMCoEPICUPressureUlcer_v4_Artifacts.zip

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** S.2b._Data_Dictionary_-_Pressure_Ulcer_4.28.16.docx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
n/a

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Number of PICU patients for whom an assessment of immobility-related pressure ulcer risk using a standardized pressure ulcer risk assessment tool was documented within 24 hours of admission.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
The numerator statement requires that the risk assessment tool is used and documented within 24 hours of admission. The denominator statement includes all patients admitted to the PICU for at least 24 hours during a monthly or quarterly reporting period.

S.6. **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, 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.
A standardized, validated pressure ulcer risk assessment tool is defined as a validated assessment tool that is applied in a standardized fashion to each patient admitted to the PICU for at least 24 hours. The assessment should be based on an immobility-related pressure ulcer risk assessment tool which has been validated for the majority of the institutions' PICU patients and the assessment should occur within the 24 hours of PICU admission.

Currently, the Braden Q is the only validated immobility-related pressure ulcer risk assessment tool available for critically ill and injured children. Other validated risk assessment tools are acceptable, if available.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
All patients admitted to the PICU for at least 24 hours during a monthly or quarterly reporting period.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Children's Health

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
n/a

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
none

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
n/a

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
n/a

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
n/a

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
n/a

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Higher score

S.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
1) Identify the target population: patients admitted to the PICU within the reporting period;

2) Evaluate the charts in the patient sample to see whether the patients meet the denominator criteria: admitted to the PICU for at least 24 hours during the reporting period;

3) Evaluate the charts that meet the denominator criteria to see whether the patients meet the numerator criteria: documentation of an assessment of immobility-related pressure ulcer risk using a standardized, validated pressure ulcer risk assessment tool within 24 hours of PICU admission; and

4) Calculate performance score by dividing the numerator by the denominator.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
n/a

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
n/a

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
If data required to compute the denominator are missing, the patient is excluded from the measure entirely. As denominator elements include admission to the PICU and duration of PICU stay, we do not anticipate that many patients who should have been included in the measure will be excluded due to missing elements. If data required to compute the numerator are missing, the patient is included in the denominator but not the numerator. In this case, the patient does not meet the measure criteria.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data : Electronic Health Record, Other, Paper Medical Records

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Other Data Source (S.23): Electronic Data Warehouse

The data source for this measure is the patient medical record. Data is collected through the Electronic Health Record (EHR) system.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility, Integrated Delivery System

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **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

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
Testing_Attachment_-_Pressure_Ulcer_7.14.16-636041761340028361.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Initial Risk Assessment for Immobility-related Pressure Ulcer within 24 hours of PICU Admission
Date of Submission: 5/13/2016
Type of Measure:

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

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☒ eMeasure (HQMF) implemented in EHRs	☒ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

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

N/A

1.3. What are the dates of the data used in testing? 01 Jan 2015 – 31 Mar 2015

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.26***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

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

The Chicago Pediatric Quality and Safety Consortium was used for measure testing. This eMeasure was tested for feasibility in four hospitals located in and around Chicago, IL and was tested for reliability in one of the hospitals:

Christ Hope Children's Hospital is a children's hospital with a pediatric residency program and fellowships in Pediatric Critical Care and Pediatric Cardiology. The hospital as 89 pediatric beds including 24 PICU beds and has approximately 6,502 pediatric admissions annually. Types of specialty care include anesthesiology, cardiology, cardiovascular surgery, neurosurgery, pulmonology, general surgery, and urology. Approximately, 35.80% of the patient population is White, 30.40% is Black, 20.10% is Other, 12.30% is unknown, 0.70% is Asian, 0.50% is Native American/Alaska Native, and 0.10% declined. More than half of patients (53.00%) use Medicaid, 45.80% use Managed Care, and 1.20% use another form of insurance.

Lutheran General Children's Hospital is a children's hospital with a pediatric residency program and fellowships in Pediatric Critical Care and Pediatric Cardiology. The hospital has 160 pediatric beds including 17 PICU beds and has approximately 7,296 pediatric admissions annually. Types of specialty care include anesthesiology, cardiology, cardiovascular surgery, neurosurgery, pulmonology, general surgery, and urology. Approximately, 44.24% of the patient population is White, 20.50% is unknown, 17.79% is Hispanic/Latino, 8.33% is Asian, 4.55% is Black, 4.22% is Other, 0.21% is Native American/Alaska Native, 0.09% declined, and 0.07% is Pacific Islander/Hawaiian. More than half of patients use Managed care (57.75%) whereas 40.81% use Medicaid, 1.20% use another form of insurance, and 0.24% are Self-pay.

Ann and Robert H. Lurie Children's Hospital is a standalone children's hospital with numerous pediatric residency and fellowship programs including programs in Neurology, Congenital Heart Surgery, Critical Care Medicine, Emergency Medicine, Pediatric Surgery, and Surgical Critical Care. The hospital has 288 pediatric beds including 40 PICU beds and has approximately 11,291 pediatric admissions annually. Types of specialty care include critical care medicine, emergency medicine, general pediatric surgery, and transplantation. Approximately, 51.80% of the patient population is White, 20.00% is Hispanic/Latino, 19.19% is Black, 4.59% is Asian, 4.59% is Other, 0.27% is unknown, and 0.27% declined. The majority of patients use either Medicaid (37.57%) or Blue Cross Blue Shield (35.95%) while 25.41% have Managed care, 0.54% have Commercial insurance, and 0.54% are insured through the government.

John H. Stroger, Jr Hospital of Cook County is the only public safety net hospital in the Chicago area. The 464-bed hospital is anchored by 228 medical/surgical beds, with dedicated units for obstetrics (40 beds), pediatrics (40 beds), intensive care (80 beds), neonatal intensive care (58 beds), and burns (18 beds). Stroger is a Level 1 Trauma Center is which treats 45,000 children and adolescents each year in the emergency room. Approximately, 55.05% of the patient population is Black, 23.01% is White , 25.09 is Hispanic/Latino, 10% is Native American 4.95% is Asian, 7% is unknown .More than half of patients (54.92%) use Medicaid, 16.65% use private insurance or self-pay, and 14.49% are charity care. The Division of Pediatric Critical Care Medicine at John H. Stroger, Jr. Hospital of Cook County in Chicago, Illinois, offers patient care in the Pediatric Intensive Care Unit at John H. Stroger Jr. Hospital of Cook County and Rush University Medical Center. The program is staffed by ten Board Certified Pediatric Intensivists with a wide range of experience, and includes on campus coverage 24/7. Members of the Stroger nursing, medical, and ancillary staff take a family-centered approach to providing the best care available to children who require intensive care services. A pediatric critical care transport is available to transport critically ill children directly to our pediatric intensive care unit. The pediatric critical care program at Stroger provides services for children with a wide range of severe illness, including the following: Trauma/burns, Severe asthma and respiratory illness, Sepsis, Cancer, Major surgery including pediatric, urology, and neurosurgery, Severe neurologic disorders, including status epilepticus, Metabolic disorders.

1.6. How many and which <u>patients</u> were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex,*

race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample)

Patients were included in the testing if they were admitted to the PICU during 01 Jan – 31 March 2015 at Lurie Children's Hospital.

Lurie Children's Hospital was able to assess this eMeasure electronically, providing output for 106 unique patients, representing 109 events.

Table 1. eMeasure Patient Characteristics	
Patient Characteristic	**N (%)**
Race/Ethnicity	
White	39 (37%)
Black	17 (16%)
Hispanic	36 (34%)
Other	13 (12%)
Unknown	1 (1%)
Insurance Status	
Private	65 (61%)
Medicaid	41 (42%)
Language Preference	
English	88 (83%)
Spanish	16 (15%)
Other	2 (2%)

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.

Four sites completed feasibility testing; however, John H Stroger Hospital currently administers the Braden Q on paper and it is not integrated into the electronic health record. As a result, while this measure is feasible at Stroger Hospital (i.e., the structured queriable fields exist in the EHR), reliability testing was not conducted at Stroger Hospital due to the fact that we would be unable to tell from the EHR fields whether a Braden Q was administered. The two Advocate Hospitals could not implement this measure into their EHR and receive reports in the timeframe required so they performed manual chart reviews (results not reported in this submission). Reliability testing of this eMeasure was conducted at Lurie Children's hospital. Demographic information on the patients included in the testing are reported above in Section 1.6.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

SDS variables included in the analysis are age, race/ethnicity, insurance status, and preferred language.

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*)

Lurie Children's Hospital implemented this measure in their EHR using an electronic algorithm which computed the measure automatically and generated a performance report on the selected patients (admitted to the PICU between 01 Jan – 31 Mar 2015). At the same time, a trained chart abstracter performed manual chart reviews on the same patients. Manual chart abstraction was then compared to the automated data abstraction to determine how reliably the overall measure and the individual measure elements were calculated.

To complete the manual abstraction when conducting parallel forms testing to assess the reliability of the eMeasure, the following algorithm was followed:

1. Evaluate the charts in the patient sample to see whether the patients meet the denominator criteria: admitted to the PICU for at least 24 hours during the reporting period;
2. Collect demographics (SDS) and elements for equity assessment: age, race/ethnicity, language preference, insurance status/type;
3. Review patient chart and document measure elements in the chart abstraction tool including both denominator and numerator measure elements; and
4. Note relevant comments.

Data analysis included inter-rater reliability (kappa).

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*)

At Lurie Children's Hospital, chart abstractions were performed for five patient charts for patient-level data included in the electronic output. Agreement for parallel-forms reliability testing was 100% for measure elements: admission date, race, ethnicity, payer, and whether a pressure ulcer risk assessment was performed within 24 hours of admission. Similarly, agreement was 100% for overall clinical performance of the measure. As agreement was 100% with no variability, a kappa statistic cannot be computed.

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?*)

As the inter-rater reliability score was 100%, the results indicate that this measure has good reliability as compared to manual chart reviews, the gold standard.

2b2. VALIDITY TESTING

2b2.1. 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*)

2b2.2. 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)*

Lurie Children's Hospital implemented this measure in their EHR using an electronic algorithm which computed the measure automatically and generated a performance report on the selected patients (admitted to the PICU between 01 Jan – 31 Mar 2015). At the same time, a trained chart abstracter performed manual chart reviews on the same patients. Manual chart abstraction was then compared to the automated data abstraction to determine how reliably the overall measure and the individual measure elements were calculated.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

As the inter-rater reliability score was 100% for all measure elements and the measure performance overall, the results indicate that this eMeasure has good validity as compared to manual chart reviews, the gold standard.

2b2.4. 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 the measure elements and measure performance scores were identical if the charts were abstracted manually or if the electronic algorithm was used, this indicates that the measure scores reflect the quality of care provided and that this eMeasure is valid.

2b3. EXCLUSIONS ANALYSIS

NA ☒ no exclusions — *skip to section 2b4*

2b3.1. 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*)

2b3.2. 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*)

2b3.3. 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*)

2b4. 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 2b5.

2b4.1. What method of controlling for differences in case mix is used?
☒ **No risk adjustment or stratification**
☐ **Statistical risk model with** Click here to enter number of factors **risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use 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.

Measure performance was tested across race/ethnicity groups, health insurance plans, patient preferred language, and age groups. The measure performed well across all groups with no statistically significant differences found in the results. Please see below for performance scores by patient factors.

Table 4. eMeasure Testing Results: Patient Factors

Patient Factors	Sub-Factors	Performance Score
Age	0 - < 6 years	92%
	6 - < 13 years	94%
	13 - < 19 years	95%
	19+ years	100%
Race/Ethnicity	White	97.5%
	Black	82%
	Hispanic	94%
	Other	92%
	Unknown	100%
Health Insurance Provider	Medicaid	90%
	Private	95%
Preferred Language	English	95%
	Spanish	88%
	Other	100%

Based on these results, we determined that it was unnecessary to control for patient factors or to stratify by patient factors when using this measure as the measure performs well across race/ethnicity groups, age groups, health insurance providers, and patients with varying preferred languages.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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)

N/A

2b4.4a. What were the statistical results of the analyses used to select risk factors?

N/A

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

N/A

2b4.5. 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 2b4.9

N/A

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

N/A

2b4.11. 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

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE
2b5.1. 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)*

Chi-square statistics were used to determine if there were statistically significant differences between age groups, race/ethnicities, health insurance plans (private vs. Medicaid), and preferred language.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

As reported above, there were no significant differences in measure performance between race/ethnicity groups (p=0.3110), age groups (p=0.8837), preferred language (p=0.5350), or insurance providers (p=0.2978).

2b5.3. 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?*)

As there were no statistically significant differences between measured entities, we can conclude that measure performance was similar across patient factors and hospitals.

2b6. 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 SDS 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 SDS 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.***

2b6.1. 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

2b6.2. 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

2b6.3. 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

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

In order to meet the denominator criteria for the measure, all components of the denominator must be present in the patient chart.

In order to meet the numerator criteria for the measure, patients must have had a pressure ulcer risk assessment performed within 24 hours of PICU admission. If data is missing, it is assumed that the care element was not provided and the patient chart does not meet numerator criteria.

2b7.2. 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*)

For the eMeasure testing, 109 events were identified and of those events, all met the denominator criteria. Only three events (2.75%) were missing information on pressure ulcer risk assessment and they were assumed to have not been performed.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

This measure performed as expected with very minimal (if any) missing data.

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

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

If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in

electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*
ALL data elements are in defined fields in electronic health records (EHRs)

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.**

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.**
Attachment **Attachment:** Pressure_Ulcer_Feasibility_Scorecard.pdf

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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
This measure was considered "feasible with workflow modifications or changes to the EHR" in only one testing site and it was feasible in the three other sites. It was determined that this measure was not yet feasible at one of the sites due to the fact that while there is a structured field indicating that the Braden-Q was administered, the tool itself is, as of the time of this study, administered on paper and not incorporated into the EHR. At this site, this often results in tests being administered without documentation of the event in the available structured field. Currently, the number of administered Braden-Q screens that are documented in a structured field in the medical record is unknown at this site.

In the event that an institution is not using the available, structured, queriable fields for the required data elements, recommendations to modify this system to enhance the feasibility of this measure include developing an integrated tool that allows consistent capture of this data element and implementing it in hospitals that currently administer the Braden-Q on paper. This will greatly increase the chances of implementation feasibility at these institutions.

This measure underwent feasibility testing in four sites and was determined to be technically feasible (i.e. the site EHR system had structured fields for all measure elements) in all four sites. The measure was also able to be implemented in three sites. One site was unable to implement the measure due to the fact that while there was a structured field available in the EHR indicating that the Braden-Q was administered, the tool was administered on paper and not incorporated into the EHR. With workflow changes, this site would have been able to implement the eMeasure.

Therefore, our testing has indicated that most EHR systems contain all the data elements needed to compute the performance measure score from electronic sources.

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.

Planned	Current Use (for current use provide URL)
Public Reporting Payment Program Professional Certification or Recognition Program Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Quality Improvement (Internal to the specific organization) Not in use	

4a.1. For each CURRENT use, checked above, provide:

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

n/a

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

This measure is not yet endorsed.

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

This measure is being submitted for endorsement for use in public and private health plans, Medicaid, and CHIPRA to assess the quality of care related to the prevention of pressure ulcers for children in the PICU for public reporting and quality improvement. This measure can also become a part of an American Board of Pediatrics (ABP) Maintenance of Certification (MOC) Performance Improvement Module (PIM). This measure will also be tested as a discrete module in the Virtual Pediatric System (VPS) pending receipt of funding from AHRQ.

4b. 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.

4b.1. Progress on Improvement. (Not required for initial endorsement unless available.)

Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

n/a

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
n/a

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
There were no unintended negative consequences to individuals or populations identified during testing.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0337 : Pressure Ulcer Rate (PDI 2)
0539 : Pressure Ulcer Prevention Implemented during Short Term Episodes of Care

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

5a. Harmonization

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 completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
NQF measure #0539, Pressure Ulcer Prevention and Care, is a pressure ulcer prevention measure targeted towards the adult population in a home health setting. While this measure appears to be somewhat comparable to the PICU measure we are proposing, our measure is designed for critically ill and injured children in the PICU, an entirely different patient population and medical care setting. NQF measure #0337, Pressure Ulcer Rate (PDI2), is a measure that captures the rate of Stage III or IV pressure ulcers in patients age 17 and younger but excludes neonates, stays less than 5 days, transfers from another facility, obstetric discharges, cases with diseases of the skin, subcutaneous tissue and breast, discharges in which debridement or pedicle graft is the

only operating room procedure, discharges with debridement or pedicle graft before or on the same days as the major operating room procedure, and discharges in which pressure ulcer is the principal diagnosis or secondary diagnosis of Stage III or IV pressure ulcer is present on admission. While this measure is targeted at the same age group as our proposed measure, the current endorsed measure assesses the percentage of patients who have a Stage III or IV pressure ulcer. Our measure requires the use of a validated tool to assess immobility pressure ulcer risk in order to prevent the occurrence of developing a pressure ulcer at all. Our measure is applied only to the care of critically ill and injured children in the PICU, a more circumscribed, but more at risk population.

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.)
No PICU-related measures are currently included in the Core Set of Children's Health Care Quality Measures for Medicaid and CHIP (Child Core Set), yet the PICU is where a hospital's sickest and most vulnerable children are treated. In addition to closing gaps in safety and/or quality, implementation of appropriate measures in the PICU could mitigate much of the elevated risk and costs associated with pediatric critical care.

The proposed measure would complement NQF measure #0337 by focusing specifically on the high-risk PICU population. We anticipate that our proposed measure will reduce the incidence of pressure ulcers in the PICU by assessing risk in a timely manner.

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:** S.25_Data_Collection_Instrument_-_Electronic_Output_-_Pressure_Ulcer_-2-.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Pediatric Consultants, LLC
Co.2 **Point of Contact:** Tom, Rice, trice@mcw.edu, 414-530-3432-
Co.3 **Measure Developer if different from Measure Steward:** Pediatric Measurement Center of Excellence
Co.4 **Point of Contact:** Ramesh, Sachdeva, rsachdeva@chw.org, 414-266-3360-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
The PMCoE Expert Work Group is composed of the following individuals:
Tom Rice (chair), Medical College of Wisconsin
Martha Curley, University of Pennsylvania School of Nursing
Daniela H Davis, University of Pennsylvania School of Medicine
Scottie B Day, Kentucky Children's Hospital, UK Healthcare
Maude Dull, Huntsville Hospital for Women and Children
Jonathon D Feldman, Kaiser Santa Clara Medical Center
Michael Forbes, Akron Children's Hospital
Hilary Franke, University of Arizona, Tucson Medical Center
Arvind K Goyal, Illinois Department of Healthcare and Family Services

Howard Jeffries, Seattle Children's Hospital
Vicki Montgomery, University of Louisville, Kosair Children's Hospital, Norton Healthcare
Michele Moss, Arkansas Children's Hospital
Matthew Niedner, University of Michigan Medical Center, Mott Children's Hospital
Gregory A Ross, Brenner Children's Hospital, Wake Forest Baptist Health
Peter Silver, Steven and Alexandra Cohen Children's Medical Center of New York, North Shore, Long Island Jewish Health System
Sophia Smith, Shady Grove Hospital & Children's National Medical Center
David C Stockwell, Children's National Medical Center
Ann E Thompson, Children's Hospital of Pittsburgh, University of Pittsburgh School of Medicine
Beth Daley Ullem, Parent Representative
Donald D Vernon, University of Utah
Derek Wheeler, Cincinnati Children's Hospital Medical Center
Lisa Wise, Parent Representative

The PMCoE Leadership Team and Staff is composed of the following individuals:
Medical College of Wisconsin, Children's Hospital and Health System: Lisa Ciesielczyk, Jaime Fox, Evelyn Kuhn, Theresa Mikhailov, Tom Rice, Ramesh Sachdeva, Matt Scanlon

American Academy of Pediatrics: Lisa Krams, Ramesh Sachdeva, Melissa Singleton, Fan Tait

Northwestern University, Feinberg School of Medicine: Lindsay DiMarco, Ray Kang, Jin-Shei Lai, Nicole Muller, Donna Woods

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:**
Ad.3 **Month and Year of most recent revision:**
Ad.4 **What is your frequency for review/update of this measure?**
Ad.5 **When is the next scheduled review/update for this measure?**

Ad.6 **Copyright statement:**
Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 3006

Measure Title: Initial Baseline Screen of Nutritional Status for Every Patient within 24 Hours of PICU Admission

Measure Steward: Pediatric Consultants, LLC

Brief Description of Measure: The measure will determine the percentage of pediatric intensive care unit (PICU) patients for whom an initial nutritional status screening was performed. The screening is to be performed within the first 24 hours of admission to the PICU with the use of a standardized nutrition-screening tool. The results of the screening must be documented in the patient's chart upon completion.

Developer Rationale: Children of all ages are at risk for malnutrition and for worsening nutritional status during a critical illness (ref.1-4). Several prospective studies and one retrospective study report the prevalence of malnutrition at admission to the PICU to range from 24% to 53% (ref.1,2,4). Further, the prevalence of weight loss during hospitalization for children ranges from 51.6% to 65% (ref.5,6). A retrospective study of critically ill children found that only 40% received any nutrition in the first 24 hours of PICU admission and caloric goals were not achieved until day 5 of PICU admission (ref.7). In hospitalized children, malnutrition is associated with an increased PICU length of stay and an increased risk-adjusted mortality rate (ref.8).

The high prevalence of malnutrition in children admitted to the PICU and the demonstration of worsening nutritional status over the course of stay in the PICU suggest that identification of nutritionally at-risk patients at the time of admission would provide an opportunity to improve nutrition therapy for these patients. An initial baseline screen of nutritional status for every patient increases provider awareness of patients' nutritional states, specifically identifying the subset of PICU patients who are at risk of malnutrition, and allows providers to adjust the timing, content, and quantity of nutrition therapy to meet the individual patient needs. Ultimately, early identification leads to early treatment which decreases PICU length of stay and mortality rates as evidenced by two recently published studies (ref.3,9).

1. Hulst J, Joosten K, Zimmermann L, Hop W, van Buuren S, Buller H,… van Goudoever J. Malnutrition in critically ill children: from admission to 6 months after discharge. Clin Nutr. 2004;23(2):223-232.

2. Hulst JM, van Goudoever JB, Zimmermann LJ, Hop WC, Albers MJ, Tibboel D, Joosten KF. The effect of cumulative energy and protein deficiency on anthropometric parameters in a pediatric ICU population. Clin Nutr. 2004;23:1381-1389.

3. Mehta NM, Bechard LJ, Cahill N, Wang M, Day A, Duggan CP, Heyland DK. Nutritional practices and their relationship to clinical outcomes in critically ill children – an international multicenter cohort study. Crit Care Med. 2012;40:2204-2211.

4. Delgado AF, Okay TS, Leone C, Nichols B, Del Negro GM. Vaz FA. Hospital malnutrition and inflammatory response in critically ill children and adolescents admitted to a tertiary intensive care unit. Clinics (Sao Paulo). 2008;63(3):357-362.

5. Sermet-Gaudelus I, Poisson-Salomon AS, Colomb V, Brusset MC, Mosser F, Berrier F, Ricour C. Simple pediatric nutritional risk score to identify children at risk of malnutrition. Clin Nutr. 2000;72:64-70.

6. Rocha GA, Edmundo, Rocha JM, Martins CV. The effects of hospitalization on the nutritional status of children. J Pediatr (Rio J). 2006;82(1):70-74.

7. de Neef M, Geukers VG, Dral A, Lindeboom R, Sauerwein HP, Bos AP. Nutritional goals, prescription, and delivery in a pediatric intensive care unit. Clin Nutr. 2008;27(1):65-71.

8. Goday PS, Kuhn EM, Sachdeva RC, Mikhailov TA. Does admission weight influence mortality and morbidiy in the Pediatric Intensive Care Unit (PICU)? JPEN J Parenter Enteral Nutr. 2008;32:316-317.

9. Goday PS, Kuhn EM, Mikhailov TA. Early parenteral nutrition is associated with significantly higher mortality in critically ill children. Presented as an oral abstract at Clinical Nutrition Week 2013. JPEN J Parenter Enteral Nutr. 2013. 37:A5-A6. Vars Candidate and Abstract of Distinction.

Numerator Statement: Number of PICU patients for whom a screening of nutritional status was documented with use of a standardized nutrition screening tool within 24 hours of admission to the PICU.

Denominator Statement: All patients admitted to the PICU for at least 24 hours during a monthly or quarterly reporting period.

Denominator Exclusions: Patients who have already had a documented nutrition screening or assessment in the previous 48 hours.

Measure Type: Process

Data Source: Electronic Clinical Data : Electronic Health Record, Other

Level of Analysis: Facility, Integrated Delivery System

New Measure - Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *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.

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 developers provide evidence based on clinical guidelines from the American Society for Parenteral and Eternal Nutrition . The guideline states “children admitted with critical illnesses should undergo nutrition screening to identify those with existing malnutrition or those who are nutritionally at-risk.”

- The developers cite a systematic review and studies published after the systematic review that demonstrate the that the majority of children present to the PICU with indices of malnutrition and that throughout PICU stay, negative energy and protein balances are common among patients and correlate with decreasing anthropometric changes.
- At the time of publication of this clinical guideline, there were no validated nutritional status screening tools in use in PICUs, and for that reason, the clinical guideline does not present estimates of benefit of nutritional screening.

Guidance from the Evidence Algorithm

1-Yes→3-Yes→4-Yes→5b→MODERATE

Questions for the Committee:

If the developer provided updated evidence for this measure:

For process 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?*

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

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

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

- This measure was tested as an eMeasure at one site, Lurie Children's Hospital. Electronic output was provided for 110 unique patients, representing 121 events. The clinical performance represented by the results of the eMeasure was good 90% of patients and 92% of screens meeting the measure.
- The eMeasure also demonstrated good clinical performance across age groups with 92% of screens performed for children 0 - <6, 96% of screens performed for children 6 - <13, and 88% of screens performed for children 13 - <19 meeting the measure. Only 67% of screens performed on patients 19 years or older met the measure due to the low sample size (N=3) in this age group.
- Reasons for not meeting the measure included not meeting the denominator criteria by having a nutrition screen more than 48 hours prior to PICU admission (N=8), not having the screen performed in the PICU (n=2), and meeting the denominator exclusion criteria by having a nutrition screen performed between 24 hours and 48 hours of PICU admission (N=5).

Disparities

- At Lurie Children's Hospital (N=105), 40% of the sample was Hispanic, 30% was White, 23% was Black, and 7% was Other. The clinical performance on the eMeasure was reasonably good across race/ethnicity groups with 97% of White patients, 88% of Black patients, 88% of Hispanic patients, and 88% of Other patients meeting the measure. These differences were not statistically significant. White patients (N=3) and Hispanic patients (N=3) were more likely than Black patients (N=0) or patients of other race/ethnicity groups (N=0) to meet the denominator exclusion criteria by already having a documented nutrition screening or assessment in the chart within 48 hours of PICU admission.

- At Lurie Children's, 54% (N=57) of the patient sample used Private Insurance and 46% (N=48) using Medicaid. Clinical performance on this eMeasure was similar in both groups with 92% of Medicaid patients and 89% of patients using Private Insurance meeting the measure. This difference was not statistically significant. Patients using private insurance were more likely to meet the denominator exclusion criteria (N=4) than Medicaid patients (N=2).

- At Lurie Children's, 77% (N=81) of the patient sample's preferred language was English as compared to 19% (N=20) who preferred Spanish and 4% (N=4) who preferred a different language. Clinical performance on this this eMeasure was good across all groups with 90% of patients who preferred English, 90% of patients who preferred Spanish, and 100% of patients who preferred a different language meeting the measure. These differences were not statistically significant. Spanish speakers were more likely to meet the denominator exclusion criteria (N=5) than English speakers (N=1) and patients who preferred a different language (N=0) and were therefore, less likely to be included in the denominator.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*

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

Comments: ** This is a process measure. Evidence is based on clinical guidelines from the Amer. Society for Parentral and Enteral Nutrition that support nutritional risk assessment. In addition, a systematic reivew of studies should that the majority of children in the PICU present with indices of malnutrition. There was no validated tool for use in the PICU

1b. Performance Gap

Comments: ** The measure testing results showed a high degree of performance in completing the screens (90% nd 92%) Performance across race/ethnicity was good, across language preferred was good and across health plans used was also good.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Electronic Clinical Data : Electronic Health Record

Specifications:

- This measure assesses the proportion of PICU patients for whom an initial nutritional status screening has been performed within 24 hours of admission.
- The measure is specified at the hospital facility or integrated delivery system level of analysis, and is meant to be reported on a monthly or quarterly basis.
- The denominator includes all patients admitted to the PICU for at least 24 hours during the reporting period.
- The numerator includes patients from the denominator population who have been assessed for risk of pressure ulcers using a standardized, validated tool.
    - The measure defines a standardized nutrition screening tool as "a screening tool that is applied in a standardized manner to each patient admitted to the PICU and should be based on a nutrition screening tool which has been validated for the majority of the institutions' PICU patients."
    - The developer notes that Examples of this would include STAMP, the Paediatric Yorkhill Malnutrition

Score, and potentially, institution-derived nutrition screening tools.

- The measure is specified as an eMeasure; a technical review of the eMeasure specifications is included below.

Questions for the Committee :

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

eMeasure Technical Advisor(s) review (if not an eMeasure, delete this section):

Submitted measure is an HQMF compliant eMeasure	The submitted eMeasure specifications follow the industry accepted format for eMeasure (HL7 Health Quality Measures Format (HQMF)). **HQMF specifications** ☒ **Yes** ☐ **No**
Documentation of HQMF or QDM limitations	N/A – All components in the measure logic of the submitted eMeasure are represented using the HQMF and QDM; **OR** **All components of the measure logic are represented correctly in the HQMF format. The developer uses User Defined QDM Value Set (1.1.1.1) for QDM data elements, but uses the topic headers of Diagnostic Test and Encounter, and the elements align with the numerator and denominator of the measures.** Submitted eMeasure contains components that cannot be represented due to limitations of HQMF or QDM and the submission explains the work around for these limitations; **OR** Submitted eMeasure contains components that cannot be represented due to limitations HQMF or QDM and the submission does NOT explain the work around for these limitations
Value Sets	The submitted eMeasure specifications uses existing value sets when possible and uses new value sets that have been vetted through the VSAC **OR** **The value sets are published within the VSAC.** Some value sets used in the submitted eMeasure are not present in the NLM Value Set Authority Center but the measure developer has provided justification for using such value sets
Measure logic is unambiguous	Submission includes test results from a simulated data set demonstrating the measure logic can be interpreted precisely and unambiguously; **OR** **The developer submitted results from both Epic and Cerner EHR systems, but acknowledged that previous screening for a nutrition exam would be in the free text sections of an EHR, which may pose issues for implementation. Nutritional status screening tools are also not standardized nor coded at this time** Submission does not include test results from a simulated data set demonstrating the measure logic can be interpreted precisely and unambiguously; **OR** Submission includes test results from a simulated data set demonstrating the measure logic cannot be interpreted precisely and unambiguously.
Feasibility Testing	The submission contains a feasibility assessment that addresses data element feasibility and follow-up with measure developer indicates that the measure logic is feasible based on

	assessment by EHR vendors**; OR** **The feasibility assessment is completed and the scores reflect the data element feasibility and the accuracy of the measure logic. However, it also does provide a current score of 2 for data availability, data standards and workflow as nutritional status screening tools are not standardized and the notation of a test is usually unstructured data.** The feasibility analysis submitted by the measure developer meets the requirements to be considered for eMeasure Trial Approval.

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☒ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☐ **Yes** ☐ **No**

Method(s) of reliability testing

- While eMeasure feasibility testing was conducted in four Chicago-area hospitals, reliability testing was only conducted in one of those hospitals because of implementation issues at the other three hospitals.
- To demonstrate reliability, the developer performed data element testing at one hospital site (Ann and Robert H. Lurie Children's Hospital) with 288 pediatric beds (including 40 PICU beds) and approximately 11,291 pediatric admissions annually.
- Patients were included in the testing if they were admitted to the PICU during 01 Jan – 31 March 2015 at Lurie Children's Hospital; the analysis included 105 unique patients, representing 121 events.
- The testing involved implementation of the eMeasure to compute scores automatically, and manual chart review of the same patients by a trained chart abstracter; inter-rater reliability was then assessed.

Results of reliability testing

- The developer reports that inter-rater reliability was conducted on five patient charts.
- Agreement was 100% for all critical data elements, and 100% for overall clinical performance of the measure.
- Because agreement was 100%, a Kappa score could not be computed.
- The developer interprets the results as indicating that the measure has good reliability when compared to the gold standard of chart reviews.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

Box 1] Specifications precise and unambiguous → [Box 2] Empirical testing NOT conducted on the measure as specified → [Box 3] Empirical validity testing of patient-level data conducted → [Box 11 of **Validity Algorithm**] Testing method described and appropriate → [Box 12] Moderate certainty or confidence that the data used in the measure are valid (Some concern about testing only being conducted in a single site and on only five patients.)

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

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Specification not completely consistent with evidence

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

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.

SUMMARY OF TESTING

Validity testing level ☐ **Measure score** ☒ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

- ☐ **Face validity only**
- ☐ **Empirical validity testing of the measure score**

Validity testing method:

- See reliability section above (data element validity testing)

Validity testing results:

- See reliability section above (data element validity testing)

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*
- *Other specific question of the validity testing?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The developer reports that the measure has one exclusion: patients who have already had a documented nutrition screening or assessment in the previous 48 hours.
- This exclusion was tested by identifying the subset of patients who met the exclusion criteria and considering patient characteristics.
- In the developer's analysis, five patients met the exclusion criteria and were not included in the measure denominator.
- The developer reports that they did not find any systematic patterns that they felt would exclude certain groups across all sites, and that because such a small number of patients meet the exclusion criteria, there is not likely to be an unfair distortion of performance results.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

N/A

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- To assess whether measure scores can detect meaningful differences in performance, the developer used Chi-square statistics to determine if there were statistically significant differences between age groups, race/ethnicities, health insurance plans (private vs. Medicaid), and preferred language.
- The developer reported that there were no significant differences in performance.
- Interpreting this information, the developer states that "measure performance was similar across measured entities."
- The developer did not analyze the measure's ability to detect meaningful differences in performance across facilities.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

N/A

2b7. Missing Data

- With regard to missing data, the developer reports that, in order to meet the denominator criteria for the measure, all components of the denominator must be present in the patient chart.
- If numerator data is missing, it is assumed that the care element was not provided and the patient chart does not meet numerator criteria.
- In data used for measure testing, 108 events were identified and five patients met the donominator exclusion criteria. Ten patients did not meet the numerator criteria either due to having a screen of nutritional status more than 48 hours prior to PICU admission (N=8) or not having the screen performed at all (N=2).
- Interpreting this analysis, the developer states that this measure performed as expected with very minimal (if any) missing data.

Guidance from the Validity Algorithm

[Box 1] Specifications consistent with evidence → [Box 2] Potential threats to validity addressed, but meaningful differences information is questionable → [Box 3] Empirical validity testing conducted using the measure as specified → [Box 6] Validity testing NOT conducted with computed measure scores → [Box 10] Validity testing conducted with patient-level data elements [Box 11] Testing method described and appropriate → [Box 12] Moderate certainty or confidence that the data used in the measure are valid (some concern about testing only being conducted in a single site)

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

Measure received a preliminary rating of "insufficient" because data were not provided to demonstrate that the measure identifies meaningful differences in performance.

Committee pre-evaluation comments

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

2a1. & 2b1. Specifications

Comments: ** Specifications are clear and meet HQMF requirements for eMeasures. Screening for nutrition was determined from free text sections of an EHR; nutritional status screening tools are not standardized or coded at this time.

**Specifications are consistent with the evidence

2a2. Reliability Testing

Comments: ** Data element reliability testing was done in 1 hospital for Jan-Mar 2015. Inter-rater reliability was 100%

2b2. Validity Testing
Comments: ** See reliability testing results

2b3. Exclusions Analysis
2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures
2b5. Identification of Statistically Significant & Meaningful Differences In Performance
2b6. Comparability of Performance Scores When More Than One Set of Specifications
2b7. Missing Data Analysis and Minimizing Bias
Comments: ** There is only one exclusion -- patients who already had a documented screening in the previous 48 hours. Using Chi-square statistics, the developer reported no significant differences in performance between age groups, race/ethnicity, health insurance plans and preferred language.

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.

- This measure is generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), Other (collected electronically using an algorithm from EHRs or an electronic data warehouse)
- Some data elements are in defined fields in electronic health records (EHRs).
- eMeasure Feasibility Scorecard included
    - The developer included feasibility scorecard results for EHR records for Epic and Cerner.
    - However, the developer did not provide OID's listed for any of the seven data elements
- The developer stated that feasibility testing was conducted on four sites and was determined to be "technically feasible, can do today" and "feasible, can do today" for implementation feasibility at three of the sites.
- For both technical feasibility and implementation feasibility, this measure was designated "feasible with workflow modifications or changes to the EHR" at one site due to two reasons.
    - The numerator element identifying whether a patient has received a nutrition screen cannot be identified in this hospital's EHR system.
    - The denominator elements, "occurrence of an administration of a nutritional status screening tool that is standardized within the institution" and the associated date, as well as the exception element, "patients who have already had a documented nutrition screening or assessment in the previous 48 hours," are captured only as free text. (In order to increase feasibility of this measure, all elements of the measure including numerator, denominator, and exception elements should be entered into structured, queriable fields as opposed to free text or associated paper forms that are scanned into the medical record.
- The developer found that Clinical Performance can be assessed through an eMeasure that will make reporting significantly less burdensome in institutions with all of the eMeasure elements in structured, queriable fields, as was true for EPIC EHR systems assessed for this measure.
- No fees, licensing, or other requirements to use any aspect of the measure as specified were identified.

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?*
- *If an eMeasure, does the eMeasure Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites?*

- *Should scanned documents qualify as being considered to be EHR records?*

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

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: ** Developer reported that feasibility testing in 4 sites was "technically feasible" Clinical performance can be assessed through an eMeasure that will make reporting less burdensome. For both technical and implementation feasibility the measure was designated feasibility with workflow modifications.

Criterion 4: Usability and Use

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

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

Current uses of the measure [from OPUS]

Publicly reported? ☐ **Yes** ☒ **No**

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

OR

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

- This measure is not currently in use, however, the developer provided a credible plan for implementation with the expected timeframes.
- This measure is being submitted for endorsement for use in public and private health plans, Medicaid, and CHIPRA to assess the quality of care related to the prevention of pressure ulcers for children in the PICU for public reporting and quality improvement.
- The developer sees this measure becoming a part of an American Board of Pediatrics (ABP) Maintenance of Certification (MOC) Performance Improvement Module (PIM).
- The developer also foresees this measure being tested as a discrete module in the Virtual Pediatric System (VPS) pending receipt of funding from AHRQ.

There were no Unexpected findings (positive or negative) during implementation

There were no unintended negative consequences to individuals or populations identified during testing.
]

***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**
4a. Accountability and Transparency *4b. Improvement* *4c. Unintended Consequences* Comments: ** This measure is not publicly reported and not used in accountability programs. The developer has submitted a plan for implementation and would be used in private and public health plans to assess quality of care

Criterion 5: Related and Competing Measures
Related or competing measures ▪ N/A **Harmonization** ▪ N/A

Pre-meeting public and member comments

•

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

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

Measure Title: Initial Baseline Screen of Nutritional Status for Every Patient within 24 Hours of PICU Admission

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

Date of Submission: 5/13/2016

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.

- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

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

Outcome

☐ Health outcome: Click here to name the health outcome

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

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

☒ Process: Screening of nutritional status

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

N/A

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

N/A

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

In critically ill children, malnutrition is associated with an increased PICU length of stay and an increased risk-adjusted mortality rate.[1] Identifying nutritionally at-risk patients as early as possible in their illness allows providers to prescribe nutrition therapy that is appropriate for patients' nutritional statuses and clinical conditions that will most effectively facilitate the healing process. An initial baseline screen of nutritional status for every patient increased awareness of the patient's nutritional states, specifically identified the subset of PICU patients who are at risk of malnutrition, and allows providers to adjust the timing, content, and quantity of nutrition therapy to meet the individual patient's needs. While there is no single, validated screening tool that is considered appropriate for critically ill and injured children, those available (including institution-derived nutrition screening tools) typically take about five minutes to administer, can be performed at the bedside, and do not generally involve a dietician.

Screening of nutrition status is fairly quick yet vitally important as the benefits of nutrition support in the critically ill patient include improved wound healing, a decreased catabolic response to injury, improved gastrointestinal structure and function, decreased PICU length of stay, and decreased mortality.[2,3] While no guidelines currently exist for specific nutrition support of critically ill children in the PICU, adult guidelines state that the initiation of nutritional support is recommended in critically ill patients.[4]

1. Goday PS, Kuhn EM, Sachdeva RC, Mikhailov TA. Does admission weight influence mortality and morbidity in the Pediatric Intensive Care Unit (PICU)? JPEN J Parenter Enteral Nutr. 2008;32:316-317.

2. Arnold M, Barbul A. Nutrition and wound healing. Plast Reconst Surg. 2006;117:42S-58S.

3. Wray CJ, Mammen JMV, Hasselgren P. Response to stress and potential benefits of nutrition support. Nutrition. 2002;18:971-977.

4. Martindale RG, McClave SA, Vanek VW, McCarthy M, Roberts P, Taylor B, … Cresci G; American College of Critical Care Medicine; A.S.P.E.N. Board of Directors. Guidelines for the provision and assessment of

nutrition support therapy in the adult critically ill patient: Society of Critical Care Medicine and American Society for Parenteral and Enteral Nutrition: Executive Summary. Crit Care Med. 2009;37(5):1757-1761.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

Mehta NM, Compher C; A.S.P.E.N. Board of Directors. A.S.P.E.N. clinical guidelines: nutrition support of the critically ill child. JPEN J Parenter Enteral Nutr. 2009;33(30):260-276.

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

On page 261, practice guideline number 1A is quoted verbatim below:

"Children admitted with critical illnesses should undergo nutrition screening to identify those with existing malnutrition or those who are nutritionally at-risk."

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

This recommendation is assigned a level of evidence of Grade D. Grade D is defined as evidence supported by level III investigations, including nonrandomized cohort studies with contemporaneous controls.

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

The grades for the guidelines are as follows:

Grade A – Supported by at least two level I investigations

Grave B – Supported by one level I investigation

Grade C – Supported by level II investigations

Grade D – Supported by level III investigations

Grade E – Supported by level IV or V evidence

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

Grading of Guidelines and Levels of Evidence were reproduced from:

Dellinger RP, Carlet JM, Masur H. Introduction. Crit Care Med. 2004;32(11)(suppl):S446.

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☒ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

N/A

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

N/A

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

N/A

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: the grading system for the evidence should be reported in section 1a.7.*)

N/A

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

N/A

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

N/A

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

N/A

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

This evidence review focused on nutrition support therapy in the PICU. The basis of this proposed measure is supported by the recommendation quoted in 1a.4.2. (Recommendation 1A) which focuses on nutrition screening.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

Recommendation 1A was based on Level III evidence, nonrandomized cohort studies with contemporaneous controls (Grade D evidence).

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

The Levels of Evidence are as follows:

Level I – Large randomized trials with clear-cut results; low risk of false positive (alpha) and/or false-negative (beta) error

Level II – Small, randomized trials with uncertain results; moderate-to-high risk of false-positive (alpha) and/or false-negative (beta) error

Level III – Nonrandomized cohort with contemporaneous controls

Level IV – Nonrandomized cohort with historical controls

Level V – Case series, uncontrolled studies, and expert opinion

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: 1976-2009

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

Three studies supported the clinical recommendation quoted above. All three studies were nonrandomized cohort studies with contemporaneous controls.

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

A cohort study conducted by Hulst et al.[1,2] of 261 children in a multidisciplinary ICU found that negative energy and protein balance correlated with decreasing anthropometric parameters. Mean energy deficits over 14 days ranged from 27 kcal/kg in preterm neonates to 12 kcal/kg in older children. Similarly, mean protein deficits ranged from 0.6 g/kg/day in preterm neonates to 0.2 g/kg/day in children. The authors noted that energy balance was calculated from estimates of RDA and not measured by indirect calorimetry due to limitations of newer methods in daily use and the lack of reference values for the younger age groups. Anthropometric methods also have limitations in critically ill children who frequently show fluid retention with resulting edema in the first days after admission. Further, the authors concluded that a 14-day monitoring period may not be adequate for measuring anthropometric changes and that a longer time period of study might have shown a stronger relationship.[1] The same children were followed-up at 6 months and almost all children had recovered their nutrition status.[2]

A different cohort study conducted by Hulst et al.[3] assessed 105 children in a multidisciplinary PICU for deficiency in serum urea, albumin, triglycerides, and magnesium within the first 24 hours after admission. Prevalence of hypomagnesemia, hypertriglyceridemia, uremia, and hypoalbuminemia were 20%, 25%, 30%, and 52%, respectively. There were no significant associations between the disorders. Ultimately, except for uremia, the authors found no significant association between abnormalities in biological parameters and changes in scores of anthropometric measurements. However, the authors note that due to the heterogeneity of their study population, it might have been difficult to find an association and that the parameters studies might be more useful in more specific diagnostic groups. The authors also did not correct for medication use, acidosis, and gastrointestinal losses which may have influenced the levels of biochemical parameters.[3]

A cohort study of the PICU conducted by Leite et al.[4] assessed anthropometry at admission and follow-up. The authors found that 65% of PICU patients presented with indices of malnutrition and chronic malnutrition was the predominant. Similarly, mortality was higher in malnourished individuals (20% vs. 12.5%) and 36% of patients showed a decrease in weight-for-height on follow up.

1. Hulst JM, van Goudoever JB, Zimmermann LJ, Hop WC, Albers MJ, Tibboel D, Joosten KF. The effect of cumulative energy and protein deficiency on anthropometric parameters in a pediatric ICU population. Clin Nutr. 2004;23(6):1381-1389.

2. Hulst J, Joosten K, Zimmermann L, Hop W, van Buuren S, Buller H, Tibboel D, van Goudoever J. Nutrition in critically ill children: from admission to 6 months after discharge. Clin Nutr. 2004;23(2):223-232.
3. Hulst JM, van Goudoever JB, Zimmermann LJ, Tibboel D, Joosten KF. The initial monitoring of routine biochemical nutritional markers in critically ill children. J Nutr Biochem. 2006;17(1):57-62.
4. Leite HP, Isatugo MK, Sawaki L, Fisberg M. Antropometric nutritional assessment of critically ill hospitalized children. Rev Paul Med. 1993;111(1):309-313.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

This clinical guideline presents studies demonstrating that the majority of children present to the PICU with indices of malnutrition and that throughout PICU stay, negative energy and protein balances are common among patients and correlate with decreasing anthropometric changes. At the time of publication of this clinical guideline, there were no validated nutritional status screening tools in use in PICUs, and for that reason, the clinical guideline does not present estimates of benefit of nutritional screening.

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

Nutritional status screening only takes a few minutes and can be performed at the patient's bedside. Most in-house screening tools do not require a dietician to be present. Identifying patients with malnutrition allows providers to address the patients' nutritional needs ultimately improving healing and reducing mortality and morbidity.

In a systematic review of studies of hospitalized children and obesity:

- 10 of 21 studies showed a positive relationship between obesity and mortality
- Studies in critically ill, oncologic or stem cell transplant, and solid organ transplant patients showed a relationship between obesity and mortality
- 5 of 11 studies showed significantly longer length of stay for obese children[1].

1. Bechard LJ, Rothpletz-Puglia P, Touger-Decker R, Duggan C, Mehta NM. Influence of obesity on clinical outcomes in hospitalized children: A systematic review. JAMA Pediatr. 2013;167(5):476-482.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

de Souza Menezes F, Leite HP, Koch Noqueira PC. Malnutrition as an independent predictor of clinical outcome in critically ill children. Nutrition. 2012;28(3):267-270.

- The objective of this study was to determine the nutritional status of a cohort of children admitted to a PICU and to assess the effect of malnutrition as an independent risk factor affecting outcome in this patient group. A total of 385 children admitted to a PICU over a two year period were assessed for nutritional status at admission and clinical outcome. Outcome variables included 30 day mortality, length of PICU stay, and length of mechanical ventilation.
- A little under half (N=175, 45.5%) of all patients were malnourished at admission. Fewer malnourished patients (N=16, 9.14%) than non-malnourished patients (N=25, 11.9%) died. Malnutrition was associated with greater length of mechanical ventilation and PICU stay but not with mortality.
- Malnutrition is still common on presentation to the PICU. It is also associated with worse outcomes but it was not associated with mortality in this patient population.
- This paper supports the conclusions of the systematic review such that nutritional status should be assessed upon admission to the PICU.

Mehta NM, Bechard LJ, Cahill N, Wang M, Day A, Duggan CP, Heyland DK. Nutritional practices and their relationship to clinical outcomes in critically ill children – an international multicenter cohort study. Crit Care Med. 2012;40:2204-2211.

- This study looked at 500 patients from 31 PICUs in 8 countries, who required mechanical ventilation longer than 48 hours in the PICU
- Over 30% of patients has severe malnutrition on admission, with body mass index z-scores >2 (13.2%) or <-2 (17.1%) on admission.
- When adjusted for nutrition days, age, severity, use of motility agent, and duration of ventilation, a higher percentage of goal energy intake via enteral nutrition was significantly associated with lower 60-day mortality. Mortality was higher in patients who receiver parenteral nutrition.

Ross PA, Newth CJ, Leung D, et al. Obesity and mortality risk in critically ill children. Pediatrics. 2016;137(3):e20152035.

- Data was obtained for 127,607 children (mortality rate 2.48%) admitted to 50 PICUs.
- Being overweight was independently associated with increased PICU mortality, after controlling for severity of illness with the PIM2 score and preexisting comorbidities. Mortality had a U-shaped distribution when classified according to weight-for-age or weight-for-height/BMI
- After controlling for hospital, age group, demographic characteristics, complex chronic conditions, noncomplex chronic conditions, and PIM2 scores, being moderately underweight or being in any overweight group was independently associated with PICU mortality.

Vermilyea S, Slicker J, El-Chammas K, Sultan M, Dasqupta M, Hoffman RG, …, Goday PS. Subjective global nutritional assessment in critically ill children. JPEN J Parenter Enteral Nutr. 2013;37(5):659-666.

- In order to determine if underweight children admitted to the PICU have a higher risk of mortality than normal-weight children, the authers prospectively evaluated the nutrition status of 150 children admitted to the PICU with the use of the subjective global nutritional assessment (SGNA) and also measured commonly used anthropometric and laboratory measurements.
- SGNA ratings of well nourished, moderately nourished, and severely malnourished demonstrated moderate to strong correlation with several standard anthropometric measurements ($p<0.05$). However, laboratory markers did not demonstrate any correlation with SGNA. Length of stay, pediatric logistic

organ dysfunction, and risk of mortality were not significantly different across groups and did not correlate with SGNA.
- An assessment tool can be used to identify well nourished, moderately nourished, and severely nourished children with fairly high accuracy.

1a.8 OTHER SOURCE OF EVIDENCE

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

N/A

1a.8.1 What process was used to identify the evidence?

N/A

1a.8.2. Provide the citation and summary for each piece of evidence.

N/A

Measure Information

This document contains the information submitted by measure developers/stewards, but is organized according to NQF's measure evaluation criteria and process. The item numbers refer to those in the submission form but may be in a slightly different order here. In general, the item numbers also reference the related criteria (e.g., item 1b.1 relates to subcriterion 1b).

Brief Measure Information
NQF #: 3006 **Corresponding Measures:** De.2. **Measure Title:** Initial Baseline Screen of Nutritional Status for Every Patient within 24 Hours of PICU Admission Co.1.1. **Measure Steward:** Pediatric Consultants, LLC De.3. **Brief Description of Measure:** The measure will determine the percentage of pediatric intensive care unit (PICU) patients for whom an initial nutritional status screening was performed. The screening is to be performed within the first 24 hours of admission to the PICU with the use of a standardized nutrition-screening tool. The results of the screening must be documented in the patient's chart upon completion. 1b.1. **Developer Rationale:** Children of all ages are at risk for malnutrition and for worsening nutritional status during a critical illness (ref.1-4). Several prospective studies and one retrospective study report the prevalence of malnutrition at admission to the PICU to range from 24% to 53% (ref.1,2,4). Further, the prevalence of weight loss during hospitalization for children ranges from 51.6% to 65% (ref.5,6). A retrospective study of critically ill children found that only 40% received any nutrition in the first 24 hours of PICU

admission and caloric goals were not achieved until day 5 of PICU admission (ref.7). In hospitalized children, malnutrition is associated with an increased PICU length of stay and an increased risk-adjusted mortality rate (ref.8).

The high prevalence of malnutrition in children admitted to the PICU and the demonstration of worsening nutritional status over the course of stay in the PICU suggest that identification of nutritionally at-risk patients at the time of admission would provide an opportunity to improve nutrition therapy for these patients. An initial baseline screen of nutritional status for every patient increases provider awareness of patients' nutritional states, specifically identifying the subset of PICU patients who are at risk of malnutrition, and allows providers to adjust the timing, content, and quantity of nutrition therapy to meet the individual patient needs. Ultimately, early identification leads to early treatment which decreases PICU length of stay and mortality rates as evidenced by two recently published studies (ref.3,9).

1. Hulst J, Joosten K, Zimmermann L, Hop W, van Buuren S, Buller H,… van Goudoever J. Malnutrition in critically ill children: from admission to 6 months after discharge. Clin Nutr. 2004;23(2):223-232.

2. Hulst JM, van Goudoever JB, Zimmermann LJ, Hop WC, Albers MJ, Tibboel D, Joosten KF. The effect of cumulative energy and protein deficiency on anthropometric parameters in a pediatric ICU population. Clin Nutr. 2004;23:1381-1389.

3. Mehta NM, Bechard LJ, Cahill N, Wang M, Day A, Duggan CP, Heyland DK. Nutritional practices and their relationship to clinical outcomes in critically ill children – an international multicenter cohort study. Crit Care Med. 2012;40:2204-2211.

4. Delgado AF, Okay TS, Leone C, Nichols B, Del Negro GM. Vaz FA. Hospital malnutrition and inflammatory response in critically ill children and adolescents admitted to a tertiary intensive care unit. Clinics (Sao Paulo). 2008;63(3):357-362.

5. Sermet-Gaudelus I, Poisson-Salomon AS, Colomb V, Brusset MC, Mosser F, Berrier F, Ricour C. Simple pediatric nutritional risk score to identify children at risk of malnutrition. Clin Nutr. 2000;72:64-70.

6. Rocha GA, Edmundo, Rocha JM, Martins CV. The effects of hospitalization on the nutritional status of children. J Pediatr (Rio J). 2006;82(1):70-74.

7. de Neef M, Geukers VG, Dral A, Lindeboom R, Sauerwein HP, Bos AP. Nutritional goals, prescription, and delivery in a pediatric intensive care unit. Clin Nutr. 2008;27(1):65-71.

8. Goday PS, Kuhn EM, Sachdeva RC, Mikhailov TA. Does admission weight influence mortality and morbidiy in the Pediatric Intensive Care Unit (PICU)? JPEN J Parenter Enteral Nutr. 2008;32:316-317.

9. Goday PS, Kuhn EM, Mikhailov TA. Early parenteral nutrition is associated with significantly higher mortality in critically ill children. Presented as an oral abstract at Clinical Nutrition Week 2013. JPEN J Parenter Enteral Nutr. 2013. 37:A5-A6. Vars Candidate and Abstract of Distinction.

S.4. **Numerator Statement:** Number of PICU patients for whom a screening of nutritional status was documented with use of a standardized nutrition screening tool within 24 hours of admission to the PICU.
S.7. **Denominator Statement:** All patients admitted to the PICU for at least 24 hours during a monthly or quarterly reporting period.
S.10. **Denominator Exclusions:** Patients who have already had a documented nutrition screening or assessment in the previous 48 hours.

De.1. **Measure Type:** Process
S.23. **Data Source:** Electronic Clinical Data : Electronic Health Record, Other
S.26. **Level of Analysis:** Facility, Integrated Delivery System

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, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

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

Evidence_Attachment_-_Nutritional_Status_5_13_16.docx

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., the benefits or improvements in quality envisioned by use of this measure)*

Children of all ages are at risk for malnutrition and for worsening nutritional status during a critical illness (ref.1-4). Several prospective studies and one retrospective study report the prevalence of malnutrition at admission to the PICU to range from 24% to 53% (ref.1,2,4). Further, the prevalence of weight loss during hospitalization for children ranges from 51.6% to 65% (ref.5,6). A retrospective study of critically ill children found that only 40% received any nutrition in the first 24 hours of PICU admission and caloric goals were not achieved until day 5 of PICU admission (ref.7). In hospitalized children, malnutrition is associated with an increased PICU length of stay and an increased risk-adjusted mortality rate (ref.8).

The high prevalence of malnutrition in children admitted to the PICU and the demonstration of worsening nutritional status over the course of stay in the PICU suggest that identification of nutritionally at-risk patients at the time of admission would provide an opportunity to improve nutrition therapy for these patients. An initial baseline screen of nutritional status for every patient increases provider awareness of patients' nutritional states, specifically identifying the subset of PICU patients who are at risk of malnutrition, and allows providers to adjust the timing, content, and quantity of nutrition therapy to meet the individual patient needs. Ultimately, early identification leads to early treatment which decreases PICU length of stay and mortality rates as evidenced by two recently published studies (ref.3,9).

1. Hulst J, Joosten K, Zimmermann L, Hop W, van Buuren S, Buller H,… van Goudoever J. Malnutrition in critically ill children: from admission to 6 months after discharge. Clin Nutr. 2004;23(2):223-232.

2. Hulst JM, van Goudoever JB, Zimmermann LJ, Hop WC, Albers MJ, Tibboel D, Joosten KF. The effect of cumulative energy and protein deficiency on anthropometric parameters in a pediatric ICU population. Clin Nutr. 2004;23:1381-1389.

3. Mehta NM, Bechard LJ, Cahill N, Wang M, Day A, Duggan CP, Heyland DK. Nutritional practices and their relationship to clinical outcomes in critically ill children – an international multicenter cohort study. Crit Care Med. 2012;40:2204-2211.

4. Delgado AF, Okay TS, Leone C, Nichols B, Del Negro GM. Vaz FA. Hospital malnutrition and inflammatory response in critically ill children and adolescents admitted to a tertiary intensive care unit. Clinics (Sao Paulo). 2008;63(3):357-362.

5. Sermet-Gaudelus I, Poisson-Salomon AS, Colomb V, Brusset MC, Mosser F, Berrier F, Ricour C. Simple pediatric nutritional risk score to identify children at risk of malnutrition. Clin Nutr. 2000;72:64-70.

6. Rocha GA, Edmundo, Rocha JM, Martins CV. The effects of hospitalization on the nutritional status of children. J Pediatr (Rio J). 2006;82(1):70-74.

7. de Neef M, Geukers VG, Dral A, Lindeboom R, Sauerwein HP, Bos AP. Nutritional goals, prescription, and delivery in a pediatric intensive care unit. Clin Nutr. 2008;27(1):65-71.

8. Goday PS, Kuhn EM, Sachdeva RC, Mikhailov TA. Does admission weight influence mortality and morbidiy in the Pediatric Intensive Care Unit (PICU)? JPEN J Parenter Enteral Nutr. 2008;32:316-317.

9. Goday PS, Kuhn EM, Mikhailov TA. Early parenteral nutrition is associated with significantly higher mortality in critically ill children. Presented as an oral abstract at Clinical Nutrition Week 2013. JPEN J Parenter Enteral Nutr. 2013. 37:A5-A6. Vars Candidate and

Abstract of Distinction.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
This measure was tested as an eMeasure at one site, Lurie Children's Hospital. Electronic output was provided for 110 unique patients, representing 121 events. The clinical performance represented by the results of the eMeasure was good 90% of patients and 92% of screens meeting the measure. The eMeasure also demonstrated good clinical performance across age groups with 92% of screens performed for children 0 - <6, 96% of screens performed for children 6 - <13, and 88% of screens performed for children 13 - <19 meeting the measure. Only 67% of screens performed on patients 19 years or older met the measure due to the low sample size (N=3) in this age group. Reasons for not meeting the measure included not meeting the denominator criteria by having a nutrition screen more than 48 hours prior to PICU admission (N=8), not having the screen performed in the PICU (n=2), and meeting the denominator exclusion criteria by having a nutrition screen performed between 24 hours and 48 hours of PICU admission (N=5).

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*
At Lurie Children's Hospital (N=105), 40% of the sample was Hispanic, 30% was White, 23% was Black, and 7% was Other. The clinical performance on the eMeasure was reasonably good across race/ethnicity groups with 97% of White patients, 88% of Black patients, 88% of Hispanic patients, and 88% of Other patients meeting the measure. These differences were not statistically significant. White patients (N=3) and Hispanic patients (N=3) were more likely than Black patients (N=0) or patients of other race/ethnicity groups (N=0) to meet the denominator exclusion criteria by already having a documented nutrition screening or assessment in the chart within 48 hours of PICU admission.

At Lurie Children's, 54% (N=57) of the patient sample used Private Insurance and 46% (N=48) using Medicaid. Clinical performance on this eMeasure was similar in both groups with 92% of Medicaid patients and 89% of patients using Private Insurance meeting the measure. This difference was not statistically significant. Patients using private insurance were more likely to meet the denominator exclusion criteria (N=4) than Medicaid patients (N=2).

At Lurie Children's, 77% (N=81) of the patient sample's preferred language was English as compared to 19% (N=20) who preferred Spanish and 4% (N=4) who preferred a different language. Clinical performance on this this eMeasure was good across all groups with 90% of patients who preferred English, 90% of patients who preferred Spanish, and 100% of patients who preferred a different language meeting the measure. These differences were not statistically significant. Spanish speakers were more likely to meet the denominator exclusion criteria (N=5) than English speakers (N=1) and patients who preferred a different language (N=0) and were therefore, less likely to be included in the denominator.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.
n/a

1c. High Priority (previously referred to as High Impact)
The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. Demonstrated high priority aspect of healthcare
A leading cause of morbidity/mortality, Patient/societal consequences of poor quality

1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
A recent international multicenter prospective study found that 30% of children admitted to the PICU were malnourished. Of these children, 17.1% were severely underweight at admission and 13.2% were severely overweight at admission. An additional 30% of children were moderately malnourished at the time of admission to the PICU with 14.4% moderately underweight and 16.3% moderately overweight (ref.1). Another prospective study found that 18.7% of hospitalized children were severely malnourished at the time of admission to the hospital and 51.6% of patients lost weight during their hospital stays. Further, children who were malnourished on admission were still malnourished at hospital discharge and 10 (9.17%) well-nourished children developed mild malnutrition while hospitalized (ref.2). In critically ill children, malnutrition is associated with increased PICU length of stay and increased risk-adjusted mortality (ref.3). The benefits of nutrition support in the critically ill patient include improved wound healing, a decreased catabolic response to injury, and improved gastrointestinal structure and function (ref.4,5). Provided the high prevalence rates of malnutrition in critically ill children and the impact that malnutrition can have on PICU length of stay and mortality rates, nutrition screening upon admission to the PICU is a high priority aspect of healthcare for pediatric patients.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
1. Mehta NM, Bechard LJ, Cahill N, Wang M, Day A, Duggan CP, Heyland DK. Nutritional practices and their relationship to clinical outcomes in critically ill children – an international multicenter cohort study. Crit Care Med. 2012;40:2204-2211.

2. Rocha GA, Edmundo, Rocha JM, Martins CV. The effects of hospitalization on the nutritional status of children. J Pediatr (Rio J). 2006;82(1):70-74.

3. Goday PS, Kuhn EM, Sachdeva RC, Mikhailov TA. Does admission weight influence mortality and morbidiy in the Pediatric Intensive Care Unit (PICU)? JPEN J Parenter Enteral Nutr. 2008;32:316-317.

4. Arnold M, Barbul A. Nutrition and wound healing. Plast Reconstr Surg. 2006;117:42S-58S.

5. Wray CJ, Mammen JMV, Hasselgren P. Response to stress and potential benefits of nutrition support. Nutrition. 2002;18:971-977.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
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 subcriteria 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):*
Prevention, Prevention : Malnutrition, Prevention : Screening

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Health and Functional Status, Health and Functional Status : Development/Wellness, Health and Functional Status : Functional Status, Prevention, Prevention : Nutrition, Prevention : Screening

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.)*
http://www.ahrq.gov/sites/default/files/wysiwyg/policymakers/chipra/factsheets/chipra-16-p002-3-ef.pdf

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 an eMeasure **Attachment:** S.2a._Measure_Specs_-_Nutritional_Status.pdf,PMCoEPICUNutritionalStatus_v4_Artifacts.zip

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** S.2b._Data_Dictionary_-_Nutritional_Status_4.28.16.docx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
n/a

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Number of PICU patients for whom a screening of nutritional status was documented with use of a standardized nutrition screening tool within 24 hours of admission to the PICU.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
The numerator statement requires that the risk assessment tool is used and documented within 24 hours of admission to the PICU. The denominator statement includes all patients admitted to the PICU for at least 24 hours during a monthly or quarterly reporting period.

S.6. **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, 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.
A standardized nutrition screening tool is a screening tool that is applied in a standardized manner to each patient admitted to the PICU and should be based on a nutrition screening tool which has been validated for the majority of the institutions' PICU patients.

Examples of this would include STAMP, the Paediatric Yorkhill Malnutrition Score, and potentially, institution-derived nutrition screening tools.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
All patients admitted to the PICU for at least 24 hours during a monthly or quarterly reporting period.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Children's Health, Populations at Risk

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
n/a

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Patients who have already had a documented nutrition screening or assessment in the previous 48 hours.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
n/a

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
n/a

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
n/a

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Higher score

S.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
1) Identify the target population: patients admitted to the PICU within the reporting period;

2) Evaluate the charts in the patient sample to see whether the patients meet the denominator criteria: patients admitted to the PICU for at least 24 hours;

3) Evaluate the charts the meet the denominator criteria for the exclusion criteria, patients who have already had a documented nutrition screening or assessment in the previous 48 hours, and remove them from the denominator population;

4) Evaluate the remaining charts to see whether they meet the numerator criteria: PICU patients for whom a screening of nutritional status was documented with the use of a standardized nutrition screening tool within 24 hours of admission; and

5) Calculate the performance score by dividing the numerator by the denominator

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
n/a

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*

IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
n/a

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
If data required to compute the denominator are missing, the patient is excluded from the measure entirely. As denominator elements include admission to the PICU and duration of PICU stay, we do not anticipate that many patients who should have been included in the measure will be excluded due to missing elements. If data required to compute the numerator are missing, the patient is included in the denominator but not the numerator. In this case, the patient does not meet the measure criteria.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data : Electronic Health Record, Other

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Other Data Source (S.23): Electronic Data Warehouse

The data source for this measure is the patient medical record. Data is collected for the construction of the measure through the Electronic Health Record (EHR) system.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility, Integrated Delivery System

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **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

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
Testing_Attachment_-_Nutritional_Status_7.14.16.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Initial Baseline Screen of Nutritional Status for Every Patient within 24 Hours of PICU Admission
Date of Submission: 5/13/2016
Type of Measure:

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

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion

impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☒ eMeasure (HQMF) implemented in EHRs	☒ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

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

N/A

1.3. What are the dates of the data used in testing? 01 Jan 2015 – 31 Dec 2015

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.26***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

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

Measure testing was conducted in the Chicago Pediatric Quality and Safety Consortium (CPQSC), which is comprised of 5 Chicago area hospitals including Roberta and Anne Lurie Children's Hospital; Lutheran General Children's Hospital; Christ Hope Children's Hospital; and John H. Stroger Cook County Hospital was used for measure testing. The advantages of the CPQSC for measure testing is that it is comprised of different types of hospital settings an urban teriary/quatrinary hospital, 2 suburban children's hospitals and an urban safety net hospital. This measure was tested for feasibility in these four hospitals and was determined to be technically feasible in all 4 hospitals and implementable in 3 hospitals. Due to time constraints, this eMeasure was only tested for reliability in one of the hospitals. :

Christ Hope Children's Hospital is a suburban children's hospital with a pediatric residency program and fellowships in Pediatric Critical Care and Pediatric Cardiology. The hospital has 89 pediatric beds including 24 PICU beds and has approximately 6,502 pediatric admissions annually. Types of specialty care include anesthesiology, cardiology, cardiovascular surgery, neurosurgery, pulmonology, general surgery, and urology. Approximately, 35.80% of the patient population is White, 30.40% is Black, 20.10% is Other, 12.30% is unknown, 0.70% is Asian, 0.50% is Native American/Alaska Native, and 0.10% declined. More than half of patients (53.00%) use Medicaid, 45.80% use Managed Care, and 1.20% use another form of insurance.

Lutheran General Children's Hospital is a suburban children's hospital with a pediatric residency program and fellowships in Pediatric Critical Care and Pediatric Cardiology. The hospital has 160 pediatric beds including 17 PICU beds and has approximately 7,296 pediatric admissions annually. Types of specialty care include anesthesiology, cardiology, cardiovascular surgery, neurosurgery, pulmonology, general surgery, and urology. Approximately, 44.24% of the patient population is White, 20.50% is unknown, 17.79% is Hispanic/Latino, 8.33% is Asian, 4.55% is Black, 4.22% is Other, 0.21% is Native American/Alaska Native, 0.09% declined, and 0.07% is Pacific Islander/Hawaiian. More than half of patients use Managed care (57.75%) whereas 40.81% use Medicaid, 1.20% use another form of insurance, and 0.24% are Self-pay.

Ann and Robert H. Lurie Children's Hospital is an inner-city standalone children's hospital with numerous pediatric residency and fellowship programs including programs in Neurology, Congenital Heart Surgery, Critical Care Medicine, Emergency Medicine, Pediatric Surgery, and Surgical Critical Care. The hospital has 288 pediatric beds including 40 PICU beds and has approximately 11,291 pediatric admissions annually. Types of specialty care include critical care medicine, emergency medicine, general pediatric surgery, and transplantation. Approximately, 51.80% of the patient population is White, 20.00% is Hispanic/Latino, 19.19% is Black, 4.59% is Asian, 4.59% is Other, 0.27% is unknown, and 0.27% declined. The majority of patients use either Medicaid (37.57%) or Blue Cross Blue Shield (35.95%) while 25.41% have Managed care, 0.54% have Commercial insurance, and 0.54% are insured through the government.

John H. Stroger, Jr Hospital of Cook County is the only public safety net hospital in the Chicago area. The 464-bed hospital is anchored by 228 medical/surgical beds, with dedicated units for obstetrics (40 beds), pediatrics (40 beds), intensive care (80 beds), neonatal intensive care (58 beds), and burns (18 beds). Stroger is a Level 1 Trauma Center is which treats 45,000 children and adolescents each year in the emergency room. Approximately, 55.05% of the patient population is Black, 23.01% is White, 25.09 is Hispanic/Latino, 10% is Native American 4.95% is Asian, 7% is unknown. More than half of patients (54.92%) use Medicaid, 16.65% use private insurance or self-pay, and 14.49% are charity care. The Division of Pediatric Critical Care Medicine at John H. Stroger, Jr. Hospital of Cook County in Chicago, Illinois, offers patient care in the Pediatric Intensive Care Unit at John H. Stroger Jr. Hospital of Cook County and Rush University Medical Center. The program is staffed by ten Board Certified Pediatric Intensivists with a wide range of experience, and includes on campus coverage 24/7. Members of the Stroger nursing, medical, and ancillary staff take a family-centered approach to providing the best care available to children who require intensive care services. A pediatric critical care transport is available to transport critically ill children directly to our pediatric intensive care unit. The pediatric critical care program at Stroger provides services for children with a wide range of severe illness, including the following: Trauma/burns, Severe asthma and respiratory illness, Sepsis, Cancer, Major surgery including pediatric, urology, and neurosurgery, Severe neurologic disorders, including status epilepticus, Metabolic disorders.

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*)

Patients were included in the reliability testing if they were admitted to the PICU during Jan 01 – March 31 2015 at Lurie Children's Hospital.

Lurie Children's Hospital was able to assess this eMeasure electronically, providing output for 105 unique patients, representing 121 events.

Table 1. eMeasure Testing Patient Characteristics	
Patient Characteristic	**N (%)**
Race/Ethnicity	
White	31 (30%)
Black	24 (23%)
Hispanic	42 (40%)
Other	8 (7%)
Unknown	0 (0%)
Insurance Status	
Private	57 (54%)
Medicaid	48 (46%)
Language Preference	
English	81 (77%)
Spanish	20 (19%)
Other	4 (4%)

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.

Four sites completed eMeasure feasibility testing; however, John H Stroger was unable to complete reliability testing for two reasons. First, the numerator element identifying whether a patient has received a nutritional screen is not specified in their EHR system in a structured, queriable field and therefore cannot be identified in this hospital's EHR system for the construction of an eMeasure. Second, the denominator elements "occurrence of an administration of a nutritional status screening tool that is standardized within the institution" and the associated date, as well as the exception element, "patients who have already had a documented nutrition screening or assessment in the previous 48 hours," are captured only as free text. In order to increase feasibility of this measure, all elements of the measure including numerator, denominator, and exception elements should be entered in structured queriable fields as opposed to free text or associated paper forms that are scanned into the medical record.

Similarly, while this eMeasure was feasible in the two Advocate hospitals, there was not enough time for these sites to implement the eMeasure and request output prior to eMeasure submission. As a result, this measure was tested for reliability in only one site, Lurie Children's Hospital.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

SDS variables included in the analysis are age, race/ethnicity, insurance status, and preferred language.

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*)

Lurie Children's Hospital implemented this measure in their EHR using an electronic algorithm which computed the measure automatically and generated a performance report on the selected patients (admitted to the PICU between 01 Jan – 31 Mar 2015). At the same time, a trained chart abstracter performed manual chart reviews on the same patients. Manual chart abstraction was then compared to the automated data abstraction to determine how reliably the overall measure and the individual measure elements were calculated.

To complete the manual abstraction when conducting parallel forms testing to assess the reliability of the eMeasure, the following algorithm was followed:

1. Evaluate the charts in the patient sample to see whether the patients meet the denominator criteria: admitted to the PICU for at least 24 hours during the reporting period;
2. Review patient chart for evidence of the exclusion element: patients who have already had a documented nutrition screening or assessment in the previous 48 hours;
3. Discard any charts that meet the exclusion criteria;
4. Collect demographics (SDS) and elements for equity assessment: age, race/ethnicity, language preference, insurance status/type;
5. Review patient chart and document measure elements in the chart abstraction tool including both denominator and numerator measure elements; and
6. Note relevant comments.

Data analysis consisted of inter-rater reliability (kappa).

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*)

At Lurie Children's Hospital, chart abstractions were performed for 5 patient charts and were compared against the same patients in the electronic output. Agreement for parallel-forms reliability testing was 100% for

measure elements: admission date, race, ethnicity, payer, and whether a nutrition screening tool was used to assess nutritional status within 24 hours of admission to the PICU. Agreement was 100% for overall measure performance. As agreement was 100% with no variability, a kappa statistic cannot be computed.

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?*)

As the inter-rater reliability score was 100%, the results of this testing indicate that this measure has good reliability as an eMeasure as compared to manual chart reviews, the gold standard.

2b2. VALIDITY TESTING

2b2.1. 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*)

2b2.2. 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)*

Lurie Children's Hospital implemented this measure in their EHR using an electronic algorithm which computed the measure automatically and generated a performance report on the selected patients (admitted to the PICU between Jan 01 – Mar 31 2015). At the same time, a trained chart abstracter performed manual chart reviews on the same patients. Manual chart abstraction was then compared to the automated data abstraction to determine how reliably the overall measure and the individual measure elements were calculated.

To complete the manual abstraction for parallel forms testing, the following algorithm was followed:

1. Evaluate the charts in the patient sample to see whether the patients meet the denominator criteria: admitted to the PICU for at least 24 hours during the reporting period;
2. Review patient chart for evidence of the exclusion element: patients who have already had a documented nutrition screening or assessment in the previous 48 hours;
3. Discard any charts that meet the exclusion criteria;
4. Collect demographics (SDS) and elements for equity assessment: age, race/ethnicity, language preference, insurance status/type;
5. Review patient chart and document measure elements in the chart abstraction tool including both denominator and numerator measure elements; and
6. Note relevant comments.

Analysis included comparing the data elements and measure performance score between the electronic output and the manual chart abstraction.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

As the values for the electronic output and the manual chart abstraction were identical, a kappa statistic cannot be computed.

2b2.4. 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 the results from the chart abstraction and the electronic output were the same, we conclude that this measure is an accurate predictor of measure performance and is a valid eMeasure.

2b3. EXCLUSIONS ANALYSIS
NA ☐ no exclusions — *skip to section 2b4*

2b3.1. 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*)

This measure has one exclusion, patients who have already had a documented nutrition screening or assessment in the previous 48 hours. This exclusion was tested by identifying the subset of patients who met the exclusion criteria and considering patient characteristics.

2b3.2. 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*)

As for the eMeasure, five patients met the exclusion criteria and therefore, were not included in the measure denominator. White patients (N=3) and Hispanic patients (N=2) were more likely than Black patients (N=0) or patients of other race/ethnicity groups (N=0) to meet the denominator exclusion criteria. Further, patients using private insurance (N=4) were slightly more likely to meet exclusion criteria than Medicaid patients (N=2). Spanish speakers (N=5) were more likely to meet the denominator exclusion criteria than other groups.

2b3.3. 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*)

We did not find any systematic patterns that we felt would exclude certain groups across all sites. Also, as such a small number of patients meet the exclusion criteria, there is not likely to be an unfair distortion of performance results.

2b4. 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 2b5.

2b4.1. What method of controlling for differences in case mix is used?
☒ **No risk adjustment or stratification**
☐ **Statistical risk model with** Click here to enter number of factors **risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use 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.

Measure performance was tested across race/ethnicity groups, health insurance plans, patient preferred language, and age groups.

Table 2. eMeasure Testing Results: Patient Factors		
Patient Factors	**Sub-Factors**	**Performance Score**
Age	0 - < 6 years	92%
	6 - < 13 years	96%
	13 - < 19 years	88%
	19+ years	67%*
Race/Ethnicity	White	97%
	Black	88%
	Hispanic	88%
	Other	88%
	Unknown	--
Health Insurance Provider	Medicaid	92%
	Private	89%
Preferred Language	English	90%
	Spanish	90%
	Other	100%

*Assumed to be due to low sample in this age group (N=3)

*Statistically significantly different from one another (p=0.009)

Based on these results, we determined that it was unnecessary to control for patient factors or to stratify by patient factors when using this measure as the measure performs well across race/ethnicity groups, age groups, health insurance providers, and patients with varying preferred languages.

2b4.3. Describe the conceptual/clinical <u>and</u> statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

N/A

2b4.4a. What were the statistical results of the analyses used to select risk factors?

N/A

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

N/A

2b4.5. 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 2b4.9

N/A

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

N/A

2b4.11. 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

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. 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)*

Chi-square statistics were used to determine if there were statistically significant differences between age groups, race/ethnicities, health insurance plans (private vs. Medicaid), and preferred language.

2b5.2. 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*)

At the one site that performed eMeasure testing, there were no significant differences in clinical performance scores for this measure between race/ethnicity groups (p=0.6854). Similarly, there were no significant differences between age groups (p=0.3374), preferred language (p=0.8647), or insurance providers (p=0.8891).

2b5.3. 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?*)

Measure performance was similar across measured entities.

2b6. 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 SDS 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 SDS 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.***

2b6.1. 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

2b6.2. 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

2b6.3. 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

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

In order to meet the denominator criteria for the measure, all components of the denominator and none of the elements of the denominator exclusion criteria may be present in the patient chart.

In order to meet the numerator criteria for the measure, patients must have had a screening of nutritional status using a standardized nutrition screening tool within 24 hours of admission. If data is missing, it is assumed that the care element was not provided and the patient chart does not meet numerator criteria.

2b7.2. 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)

For eMeasure testing, 105 patients met the denominator criteria and five patients met the denominator exclusion criteria. Ten patients did not meet the numerator criteria either due to having a screen of nutritional status more than 48 hours prior to PICU admission (N=8) or not having the screen performed at all (N=2).

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i*.e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

This measure performed as expected with very minimal (if any) missing data.

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

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

If other: collected electronically using an algorithm from EHRs or an electronic data warehouse

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)*

Some data elements are in defined fields in 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.**

This measure underwent feasibility testing in four sites and was determined to be "technically feasible, can do today" and "feasible, can do today" for implementation feasibility at three of the sites.

For both technical feasibility and implementation feasibility, this measure was designated "feasible with workflow modifications or changes to the EHR" at one site due to two reasons. First, the numerator element identifying whether a patient has received a nutrition screen cannot be identified in this hospital's EHR system. Second, the denominator elements, "occurrence of an administration of a nutritional status screening tool that is standardized within the institution" and the associated date, as well as the exception element, "patients who have already had a documented nutrition screening or assessment in the previous 48 hours," are captured only as free text. In order to increase feasibility of this measure, all elements of the measure including numerator, denominator, and exception elements should be entered into structured, queriable fields as opposed to free text or associated paper forms that are scanned into the medical record.

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.**

Attachment **Attachment:** Nutrition_Status_Feasibility_Scorecard.pdf

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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

Clinical Performance can be assessed through an eMeasure that will make reporting significantly less burdensome in institutions with all of the eMeasure elements in structured, queriable fields. This was true for EPIC EHR systems assessed for this measure. In addition, anecdotally it appeared that hospital based EHR system are more developed than office based systems, such as Cerner, and more amenable to eMeasure use.

There were two reasons that this measure was determined to be "technically feasible with workflow modifications or changes to the

EHR at John H. Stroger Jr Hospital of Cook County. First, the numerator element identifying whether a patient has received a nutritional screen cannot be identified through structured, queriable fields in this hospital's EHR system. Second, the denominator elements, "occurrence of an administration of a nutritional status screening tool that is standardized within the institution" and the associated date, as well as the exception element, "patients who have already had a documented nutrition screening or assessment in the previous 48 hours," are captured only as free text. In order to increase feasibility of this measure, all elements of the measure including numerator, denominator, and exception elements should be entered in structured, queriable fields as opposed to free text or associated paper forms that are scanned into the medical record.

This measure was designated "technically feasible with workflow modifications or changes to the EHR" for implementation feasibility because John H. Stroger Jr Hospital of Cook County does not currently administer a nutritional status screening tool. In order for this measure to be implemented at this site, the tool would need to be designed/chosen, implemented, and the staff would need to be trained to administer the tool. Additionally, discrete fields would need to exist in the EMR for required data.

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.

Planned	Current Use (for current use provide URL)
Public Reporting Payment Program Professional Certification or Recognition Program Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Quality Improvement (Internal to the specific organization) Not in use	

4a.1. **For each CURRENT use, checked above, provide:**

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

n/a

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

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
This measure is being submitted for endorsement for use in public and private health plans, Medicaid, and CHIPRA to assess the quality of care related to the prevention of pressure ulcers for children in the PICU for public reporting and quality improvement. This measure can also become a part of an American Board of Pediatrics (ABP) Maintenance of Certification (MOC) Performance Improvement Module (PIM). This measure will also be implemented in Virtual Pediatrics Systems (VPS) pending funding from AHRQ.

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

n/a

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
n/a

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
There were no unintended negative consequences to individuals or populations identified during testing.

5. Comparison to Related or Competing Measures

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

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

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 completely harmonized?

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

5b. Competing Measures

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

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

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** S.25_Data_Collection_Instrument_-_Electronic_Output_PICU_Nutritional_Status.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Pediatric Consultants, LLC
Co.2 **Point of Contact:** Tom, Rice, trice@mcw.edu, 414-530-3432-
Co.3 **Measure Developer if different from Measure Steward:** Pediatric Measurement Center of Excellence
Co.4 **Point of Contact:** Ramesh, Sachdeva, rsachdeva@chw.edu, 414-266-2000-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
The PMCoE Expert Work Group is composed of the following individuals:
Tom Rice (chair), Medical College of Wisconsin
Martha Curley, University of Pennsylvania School of Nursing
Daniela H Davis, University of Pennsylvania School of Medicine
Scottie B Day, Kentucky Children's Hospital, UK Healthcare
Maude Dull, Huntsville Hospital for Women and Children
Jonathon D Feldman, Kaiser Santa Clara Medical Center
Michael Forbes, Akron Children's Hospital
Hilary Franke, University of Arizona, Tucson Medical Center
Arvind K Goyal, Illinois Department of Healthcare and Family Services
Howard Jeffries, Seattle Children's Hospital
Vicki Montgomery, University of Louisville, Kosair Children's Hospital, Norton Healthcare

Michele Moss, Arkansas Children's Hospital
Matthew Niedner, University of Michigan Medical Center, Mott Children's Hospital
Gregory A Ross, Brenner Children's Hospital, Wake Forest Baptist Health
Peter Silver, Steven and Alexandra Cohen Children's Medical Center of New York, North Shore, Long Island Jewish Health System
Sophia Smith, Shady Grove Hospital & Children's National Medical Center
David C Stockwell, Children's National Medical Center
Ann E Thompson, Children's Hospital of Pittsburgh, University of Pittsburgh School of Medicine
Beth Daley Ullem, Parent Representative
Donald D Vernon, University of Utah
Derek Wheeler, Cincinnati Children's Hospital Medical Center
Lisa Wise, Parent Representative

The PMCoE Leadership Team and Staff is composed of the following individuals:
Medical College of Wisconsin, Children's Hospital and Health System: Lisa Ciesielczyk, Jaime Fox, Evelyn Kuhn, Theresa Mikhailov, Tom Rice, Ramesh Sachdeva, Matt Scanlon

American Academy of Pediatrics: Lisa Krams, Melissa Singleton, Fan Tait

Northwestern University, Feinberg School of Medicine: Lindsay DiMarco, Ray Kang, Jin-Shei Lai, Nicole Muller, Donna Woods

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:**
Ad.3 **Month and Year of most recent revision:**
Ad.4 **What is your frequency for review/update of this measure?**
Ad.5 **When is the next scheduled review/update for this measure?**

Ad.6 **Copyright statement:**
Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

Data Dictionary for Initial Baseline Screen of Nutritional Status for Every Patient Within 24 Hours of PICU Admission

Measure Element	**Description**	**Variable**
Initial population	# All patients discharged from the PICU during the reporting period	Intersection of: • "Encounter, Performed: PICU Admission or Transfer (Occurrence A" <= 24 hour(s) during "Measurement Period" • "Encounter, Performed: PICU Admission or Transfer (Occurrence A) (admission datetime)" • "Encounter, Performed: PICU Admission or Transfer (Occurrence A) (discharge datetime)"
Denominator	# All patients discharged from the PICU during the reporting period	Intersection of: • "Encounter, Performed: PICU Admission or Transfer (Occurrence A" <= 24 hour(s) during "Measurement Period" • "Encounter, Performed: PICU Admission or Transfer (Occurrence A) (admission datetime)" • "Encounter, Performed: PICU Admission or Transfer (Occurrence A) (discharge datetime)"
Denominator Exclusions	# Patients who have already had a documented nutrition screening or assessment in the previous 48 hours	Union of: • "Occurrence of Diagnostic Study, Performed: an administration of a nutritional status screening tool that is standardized within the institution (Occurrence C)" <= 48 hour(s) during "Encounter, Performed: PICU Admission or Transfer (Occurrence A)"
Numerator	# Number of PICU patients for whom a screening of nutritional status was documented with use of a standardized nutrition screening tool within 24 hours of admission	Intersection of: • "Occurrence of Diagnostic Study, Performed: an administration of a nutritional status screening tool that is standardized within the institution (Occurrence C)" • "Occurrence of Diagnostic Study, Performed: an administration of a nutritional status screening tool that is standardized within the institution (Occurrence C) (start datetime)"
Numerator Exclusions	None	N/A
Denominator	None	N/A

Exceptions		
Stratification	None	N/A

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 3025
De.2. **Measure Title:** Ambulatory Breast Procedure Surgical Site Infection (SSI) Outcome Measure
Co.1.1. **Measure Steward:** Surveillance Branch, Division of Healthcare Quality Promotion, Centers for Disease Control and Prevention
De.3. **Brief Description of Measure:** This measure is for the risk-adjusted Standardized Infection Ratio (SIR) for all Surgical Site Infections (SSI) following breast procedures conducted at ambulatory surgery centers (ASCs) among adult patients (ages 18 - 108 years) and reported to the Centers for Disease Control and Prevention (CDC) National Healthcare Safety Network (NHSN). The measure compares the reported number of surgical site infections observed at an ASC with a predicted value based on nationally aggregated data. The measure was developed collaboratively by the CDC, the Ambulatory Surgery Center Quality Collaboration (ASC QC), and the Colorado Department of Public Health and Environment. CDC is the measure steward.
1b.1. **Developer Rationale:** The measure provides summary results that ASCs can use as quantitative aids in their efforts to evaluate and reduce breast surgery surgical site infection rates. The SIRs can be used by ASCs to benchmark SSI rates, identify opportunities for improvement, and gauge the impact of prevention efforts. At the outset, the SIRs provide a set of signals that often warrant further analysis, such as an examination of lapses in infection control practices that may contribute to high incidence of SSI. Some of the analytic follow up can be completed with data reported to CDC's National Healthcare Safety Network (NHSN)Patient Safety Component Procedure-Associated (PA) Module, using analytic features built into the NHSN application. However, additional analyses to determine the cause of infections as targets for prevention in individual instances are likely to require access to data that is beyond the scope of data collection and analysis using the NHSN module.

Breast procedures were specifically chosen for this measure due to the observed burden of breast procedure-associated SSI. Out of 67,150 ASC procedures reported to NHSN from 2010-2013, 30,787 (45.9%) were breast procedures. Out of the 142 SSIs reported from ASCs during the same time period, 78 (54.9%) were related to breast procedures, indicating an SSI risk of 0.25%. This was the highest volume and SSI risk out of all outpatient ASC procedures reported in the timeframe.

S.4. **Numerator Statement:** Surgical site infections (SSIs) during the 30-day (superficial SSI) and 90-day (deep and organ/space SSI) postoperative periods following breast procedures in Ambulatory Surgery Centers.
S.7. **Denominator Statement:** Breast procedures, as specified by the operative codes that comprise the breast procedure category of the NHSN Patient Safety Component Protocol, performed at ambulatory surgery centers.
S.10. **Denominator Exclusions:** Hospital inpatients and hospital outpatient department patients, pediatric patients and very elderly patients, and brain-dead patients whose organs are being removed for donor purposes

De.1. **Measure Type:** Outcome
S.23. **Data Source:** Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Paper Medical Records
S.26. **Level of Analysis:** Facility

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported

by the stated rationale.

Summary of evidence:

- The overall body of evidence on the incidence, outcomes, and prevention of SSIs in the ambulatory surgical center (ASC) patient population is sparse but the available data suggest risks for SSIs following some breast procedures in some settings may be as high as 30%.In the current literature, the rates of SSI in ambulatory surgery centers is relatively low—however, aggregate numbers of infections can still cause a substantial burden, as those often result in post-surgical visits and morbidity.
- ASCs have been shown to have a lower SSI rate than inpatient settings. Though estimates of risk for breast procedures specifically vary from 1% to over 30% (and rate varies from 3 SSI to 28 SSI per 1000 procedures) depending on breast procedure type, sample population, and definition of SSI, it is clear that breast procedure-related SSIs are a large burden to outpatient healthcare facilities, and provide much room for benefit. There is little data on the number or proportion of preventable SSI specifically following breast procedures conducted in ASCs.

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*

Preliminary rating for evidence: ☒ Pass ☐ No Pass

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

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

- The developer summarized an exploratory analysis of NHSN data that showed that out of 67,150 ambulatory surgical center (ASC) procedures reported to NHSN from 2010-2013, 30,787 (45.9%) were breast procedures.
- Out of the 142 SSIs reported from ASCs during the same time period, 78 (54.9%) were related to breast procedures, indicating a risk of SSI of 0.25%. This was the highest volume and SSI risk among all outpatient ASC procedures reported in the timeframe.
- Numerous individual studies and systematic reviews provide strong evidence that measurement and feedback of surgical site infections leads to lower SSI rates in the long term.

Disparities

- Data on disparities in surgical site infections in ASCs, as well as in hospitals, are sparse. No studies or reviews were found specifically on disparities surrounding SSI in any healthcare facility. However, it has been extensively documented that surgical site infections lead to an excess cost burden as well as excess hospital stay for patients. These additional costs may cause disparities in care for SSI, which are reflective of disparities in access to health care in general.

Questions for the Committee:

- *Specific question on information provided for gap in care.*
- *Is there a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*

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
Comments: ** This is a health outcome and supported by the rationale.
** While the overall body of evidence on incidence, outcomes and prevention of SSIs in the ASC pt population is limited, the

developers discussed that the risk of SSI after breast procedure may be as high as 33%. The evidence is emerging and this measure may contribute to advance the evidence.
** Yes. Better infection control is always a possible.
** Yes, this measures a health outcome. The overall body of evidence on the incidence, outcomes, and prevention of SSIs in the ASC patient population is sparse but the available data suggest risks for SSIs following some breast procedures in some settings may be as high as 30%. Studies have shown that measurement and feedback of SSI leads to lower SSI rates in the long term.
** This is an outcome measure. Sparse data is available, but suggests that risks for SSIs following some breast procedures in some settings is as high as 30%. Current literature suggests the rate of infection in ASC is relatively low, but the burden can be significant.

1b. Performance Gap

Comments: ** Yes-Of ambulatory surgical procedures breast surgery infections account for 54% of all ambulatory surgical infections.
** Exploratory analysis of the NHSN data showed that breast procedures comprised 45.9% of the ASC procedures 2010-2013. Of the SSIs during that time, 54.9% were following breast procedure, indicating risk of 0.25%
** Wide gap apparently 1-30%. Major opportunity for improvement.
** Yes, the performance data provided demonstrates a gap in care. Out of 67,150 ASC procedures reported to NHSN from 2010-2013, 45.9% were breast procedures. Out of the 142 SSIs reported from ASCs 54.9% were related to breast procedures, indicating a risk of SSI of 0.25%. Is the volume of SSIs or related morbidity and cost sufficient to warrant a national performance measure?
** NHSN dta of 67,150 ASC procedures from 2010-13 -- 30,787 breast procedures. Of 142 SSis reported, 54.9% were related to breast procedures, indicating a risk of SSI of 0.25% -- the highest volume and SSI risk among all ASC procedures.
No data available on dispartities

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Electronic clinical data/electronic health record

Specifications:

- This measure calculates a Standardized Infection Ratio (SIR) for Surgical Site Infections (SSI) following breast procedures conducted at ambulatory surgery centers (ASCs) among adult patients (ages 18 - 108 years)
- The measure is reported as an observed-to-expected ratio, which compares the reported number of surgical infections observed at an ASC with a predicted value based on nationally-aggregated data.
- The level of analysis is ambulatory surgery center (ASC) facilities.
- The data source is the National Healthcare Safety Network (NHSN), which will collect data on SSIs following outpatient operative procedures through the new Outpatient Procedure Component in 2018.
- Regarding the time period for data, the developer states that, for NHSN purposes, data will be aggregated quarterly and annually based on the calendar year. However, facilities or groups may choose to aggregate data at different intervals (monthly, fiscal year, etc.) for their own quality initiatives.
- The denominator identifies breast procedures using CPT codes; a detailed list of codes is provided in the developer's supplemental materials.
- The numerator identifies all surgical site infections in cases included in the denominator; the developer provides definitions and criteria for surgical site infections, superficial incisional SSI, deep incisional SSI, organ/space SSI, and breast abscess/infection.
- The measure excludes hospital inpatients and hospital outpatient department patients, pediatric patients and very elderly patients, and brain-dead patients whose organs are being removed for donor purposes.

Questions for the Committee :

- *Are all the data elements clearly defined? Are all appropriate codes included?*

- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☒ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing

- The developer assessed data element reliability on procedures reported from selected ASCs in Colorado from January to December 2014.
- The distribution of facilities included in the reliability analysis by city is provided by the developer.
- A total of 18 ASCs were included in the study; Selected ASCs had performed at least 100 breast surgeries in 2014.
- A total of 715 charts were examined (701 female and 14 male) by staff from the Colorado Department of Public Health and Environment to identify under- and over-reported events, data discrepancies, and omissions in events and procedures.

Results of reliability testing

- The developer reports that no under-reported events were found and one over-reported event was identified because the case did not meet all NHSN criteria for superficial SSI.
- With respect to data discrepancies and omissions, the developer summarizes a number of instances of incorrect reporting and data entry failures or errors.
- The developer provides a table of NHSN-reported procedures and events, number of charts reviewed, ineligible procedures, and under- and over-reported events for each ASC included in the testing; a table of discrepant variables by facility is also provided.
- Data provided by the developer shows that almost all facilities reported procedure duration incorrectly because of an outdated protocol definition; however, the developer reports that this variable was not included in the final model.
- The developer's interpretation of their reliability assessment is that the measure is highly reliable, in being able to consistently and correctly identify SSI across facilities and raters. The number of data entry errors in ASA Score is relatively low (7%), and the number of errors in date of birth (Age) is very low (<1%), indicating that a performance score calculated using these elements would be reliable as well.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Is this an appropriate method for demonstrating measure reliability?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

Guidance from the Reliability Algorithm

[Box 1] Specifications precise and unambiguous → [Box 2] Empirical testing NOT conducted on the measure as specified → [Box 3] Empirical validity testing of patient-level data conducted → [Box 11 of **Validity Algorithm**] Testing method insufficient according to NQF guidance (only percent agreement provided) → [Insufficient]

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

Rationale: NQF guidance suggests that data element validity testing that provides only percent agreement, with no

additional analyses (e.g., Kappa statistic, sensitivity/sensitivity. PPV/NPV, etc.), should be rated insufficient.

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

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.

SUMMARY OF TESTING

Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

- ☒ **Face validity only**
- ☐ **Empirical validity testing of the measure score**

Validity testing method:

- To demonstrate validity of the measure score, the developer conducted a face validity assessment using a formal consensus process.
- 11 individuals working in ambulatory surgery centers (in various roles) were administered a questionnaire related to the validity, feasibility, interpretability, and actionability of the measure.
- The questionnaire rated the respondent's level of agreement with statements related to each measure attribute based on a 5-point Likert Scale with a rating of 5 expressing agreement and 1 expressing disagreement.

Validity testing results:

- The developer reports that there was high level of agreement among the respondents regarding the validity of the measure, with 9/11 (81.8%) agreeing that the measure appears to measure what it is intended to, giving a 5/5 rating response.
- 9/11 (81.8%) respondents also gave a 5/5 rating on whether the measure allows for consistent interpretation across centers, and one gave a 4/5 rating.
- 8/11 respondents agreed (with a 4/5 or 5/5 rating) that the measure's score accurately reflects the quality of a center's performance; 2 neither agreed nor disagreed (3/5); and 1 disagreed (1/5).
- Regarding the statement that the measure's score can be used to distinguish between good and poor performance, 7 respondents (63.6%) agreed, giving a minimum rating of 4/5, 3 (27.3%) gave a rating of 3/5, and 1 disagreed with the statement (1/5).
- The developer's interpretation of these results is that there was high level of agreement among the respondents regarding the validity of the measure.

Questions for the Committee:

- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- Procedures are excluded from this measure if patients are under 18 or greater than 109 years old.
- The developer limited the population to adult patients given the nature of breast surgeries, and states that ages

entered above 109 were considered data entry errors.
- The developer states that all exclusions were necessary to achieve the most accurate and applicable model.
- No statistical testing was performed on exclusions.

Questions for the Committee:
- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☐ **None** ☒ **Statistical model** ☐ **Stratification**

Conceptual rationale for SDS factors included ? ☐ **Yes** ☒ **No**

SDS factors included in risk model? ☐ **Yes** ☒ **No**

Risk adjustment summary

- The measure is risk adjusted using a statistical model with two factors: categorical ASA classification, and ordinal age categories.
- The developer notes that potential adjustment factors were limited by the scope of variables collected by NHSN. Those considered, based on factors identified in literature, were: age of patient, anesthesia use, ASA classification, duration of procedure, gender of patient, and surgical wound classification.
- Univariate analyses were conducted between each of these factors and the outcome; those showing statistically-significant associations were included in the modeling process.
- The modeling process involved a backwards elimination of predictors from the saturated model, where the least significant predictor was removed until all remaining factors were significant.
- The developer notes that duration of procedure was a significant factor, but was excluded because of clinical concerns about eligibility as a confounding factor.
- Details of the risk model are provided in a table in the submission form.
- To validate the model, the developer applied a logistical regression model with bootstrapping methods to 100 independent samples; validation estimates and corresponding 95% confidence intervals are provided in a table in the submission form.
- The developer provides discrimination statistics to demonstrate the model's ability to correctly predict outcomes in observation data (**c-index = 0.675**), and calibration statistics to demonstrate the agreement between observed outcomes and outcomes predicted by the model (**Hosmer-Lemeshow p=0.6626**).
- The developer's interpretation of these results is that the model can control for differences in patient case-mix adequately. Further measure maintenance may be required in the future to update the model with more informed and complete datasets.

Questions for the Committee:
- *Is an appropriate risk-adjustment strategy included in the measure?*
- *Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented?*
- *Are all of the risk adjustment variables present at the start of care? If not, describe the rationale provided.*
- *Do you agree with the developer's decision to not include SDS factors in their risk-adjustment model?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- With regard to the measure's ability to detect meaningful differences in performance, the developer notes that a meaningful difference in the Standardized Infection Ratio (SIR) was defined as an SIR and a confidence interval that was statistically different from 1.
- The developer notes that the SIR is not calculated when a facility's predicted value is less than 0.2.
- The developer reports that out of 138 total facilities reporting from 2010-2014, SIRs were able to be calculated for 70 of them.

- The developer provides a table showing the percentage of SIRs that were significantly different from 1.
- The developer's interpretation of their analysis is that the SIR enables detection of statistically significant and clinically meaningful differences in SSI that warrant further analysis and possible action.
- The developer suggests that, although exposure volume is low, leading to few statistically significant SIRs in this population, the value of the calculated SIRs can reflect practical measures of performance.

Question for the Committee:

○ *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

N/A

2b7. Missing Data

- To identify the extent and distribution of missing data, the developer compared the crude risks of SSI between the procedures with missing ASA class and procedures with complete ASA class using a chi-square test.
- The developer reports that there were 8,345 missing ASA Classifications out of 46,018 eligible procedures (18.13%). The crude risk in the missing procedures was not significantly different from the crude risk in the included procedures (0.36% compared to 0.25%, p=0.0714).
- The developer suggests that, based on the results above, the missing population does not seem to be significantly different from the included population, minimizing the amount of systemic bias.

Guidance from the Validity Algorithm

[Box 1] Specifications consistent with evidence → [Box 2] Potential threats to validity addressed → [Box 3] Empirical validity testing NOT conducted using the measure as specified → [Box 4] Face validity systematically assessed → [Box 5] Substantial agreement that the performance measure score can be used to distinguish quality → [Box 8b]

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

Committee pre-evaluation comments

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

2a1. & 2b1. Specifications

Comments: ** Procedures identified with CPT codes. Clear definition of criteria for infection superficial as well as deep. Incorrect reporting and data entry failures were reported. Some reporting issues attributed to using an outdated protocol which has been replaced

** Face validity only using a formal consensus process – 11 individuals working in ambulation surgery centers in various roles were administered a questionnaire related to the validity, feasibility interpretability, and action ability of the measures, with scores on each variable 5 (agree) to 1(disagree). The developers reported high level of agreement for validity (81.8%); 81.8% for consistency of interpretation across centers 73% agreement that the measure score accurately reflects quality of the center's performance; 63.6% agreed that the measure score could distinguish between good vs. poor performance

** All clear here re specifications. Should be amenable to consistent reporting. Good linkage between target population values and results reporting.

** The specifications are clearly outlined, and CPT codes are included.

** Specifications are clear and consistent with evidence.

2a2. Reliability Testing

Comments: ** Testing done at the data element with only percent agreement.

Insufficient

** Reliability: Data reliability on procedures was reported from selected ASCs in Colorado from Jan-Dec 2014 – 18 centers. 100 breast procedures were performed, with 715 pt records examined by the Colorado Dept of Health and Environment to identify under or over reporting and omissions in events and procedures. No under reports were found, 1 over report (did not meet all NHSN criteria for superficial SSI)

** Reliability approach seems appropriate. I was concerned about the small number of SSIs in the ASCs included in this aspect of the study. But this sample is used for reliability,not for overall administration of the measure, so I think it is OK.

** The developer assessed data element reliability on procedures reported from 18 ASCs in Colorado from January to December 2014. 715 patient charts were abstracted and 5 SSIs identified. No under-reported events were found and 1 over-reported event was identified. Is the sample size sufficient for generalization?
** Data element reliability testing was performed with data source an level of analysis indicated -- selected ASCs in Colorado from Jan. to Dec. 2014 -- a total of 18 ASCs. No under-reported events were found. Developer indicates measure is highly reliable in being able to consistently and correctly identify SSI across facilities and raters

2b2. Validity Testing
Comments: ** Validity testing done with face validity.
** Validity: the measure was risk adjusted using a statistical model with 2 factors: categorical ASA classification and ordinal age categories – least significant factors were removed until all remaining factors were significant. However, duration of the procedure was a significant factor, but was excluded from the model because of clinical concerns about eligibility as a confounding factor. Meaningful difference in the SIR (Standard Infection Ratio) was defined and a confidence interval that was statistically different than 1. The SIR enables detection of statistically significant and clinically meaningful differences in SSI that warrant further analysis and possible action. The extent of missing data was analyzed and the crude rate of SSI did not seem to differ from the missing classifications – minimizing bias.
** Looks OK
** The developer conducted a face validity assessment using a formal consensus process with 11 individuals working in ASCs (7 nurses, 4 surgeons). There was general agreement among the respondents regarding the validity of the measure. Comments include factors outside a facility's control (patient comorbidities, poor hygiene, and non-compliance) may affect the measure's score.
** Measure score validity testing was done using face validity through a formal consensus process -- 11 individuals from ASCs received a questionnaire. High level of agreement reported.

2b3. Exclusions Analysis
2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures
2b5. Identification of Statistically Significant & Meaningful Differences In Performance
2b6. Comparability of Performance Scores When More Than One Set of Specifications
2b7. Missing Data Analysis and Minimizing Bias

Comments: ** Modelling used backward elimination of predictors from saturated model. To validate applied logistic regression model with bootstrapping to 100 independent samples. Provided discriminator statistics to demonstrate ability to correctly predict outcome. C index 0.675. Callibration statistics to demonstrate observed and predicted Hosmer - Lemeshow p=0.0714 suggesting no significant difference.
Moderate
** No real threats
** Exclusions are supported by evidence.
A statistical model of risk adjustment is included Factors identified were age, anesthesia used, ASA classification, duration of procedure gender and surgical wound classification. Developer provides discrimination statistics that show a c-index of 0.675 and calibration statistics of 0.6626 which show the model can control for differences in patient case-mix adequately.
**Missing data (18.13%) does not seem to include a population that is different from the included population.

Criterion 3. Feasibility

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.

- Data for this measure is generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score) and abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)
- Some data elements are in defined fields in a combination of electronic sources.
- The developer states that use of NHSN surveillance protocol, definitions, and data collection methods for SSI have proven feasible across multiple healthcare settings, including ambulatory surgery centers. Facilities are instructed to follow a standardized data collection procedure (specified by the NHSN protocol and definitions),

but specific data collection methods may vary between facilities. Denominator data for breast SSIs are reported as the total number of breast surgical procedures conducted, i.e., an 100% sample. Patient-level data is reported for each procedure and infection; however, the medium of reporting through NHSN is secure and the risk of breaches in patient confidentiality is low. Technical guidance provided by CDC will aid and facilitate accurate data collection and reporting.

- No fees or licensing requirements. To use this SIR measure, ASCs must be enrolled in NHSN.

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?*
- *If an eMeasure, does the eMeasure Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites?*

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

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: ** Elements routinely used during provision of care. Some fields in electronic sources. NHSN is operational. Moderate

** Feasibility: data for this measure is generated or collected and used by healthcare personnel during the provision of care and abstracted from someone other than the person obtaining the original information. Some data elements are defined fields in the EMR. The NHSN surveillance protocols, definitions and data collection methods for SSI have proven feasible across healthcare settings. To use SIR, ASCs must enroll in NHSN

** High feasibility

** NHSN surveillance protocol, definitions, and data collection methods for SSI have proven feasible across multiple healthcare settings, including ASCs. Are all ASCs enrolled in NHSN and reporting consistently?

** Data is generated by and used by personnel in provision of care. Some data elements are in electronic sources. NHSN data are collected using a protocol and definitions

Criterion 4: Usability and Use

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

Current uses of the measure

Publicly reported? ☒ Yes ☐ No

Current use in an accountability program? ☒ Yes ☐ No

OR

Planned use in an accountability program? ☒ Yes ☐ No

Accountability program details:

- National Healthcare Safety Network (NHSN), Centers for Disease Control and Prevention. NHSN is a national system used by CDC and its partners in clinical care and public health for surveillance of healthcare-associated

infections, healthcare worker safety, blood safety, antimicrobial use and resistance, and adherence to prevention practices. The system is designed to provide actionable data for healthcare facilities and systems, public health agencies at the state and federal levels, and prevention collaborations. NHSN is the data source for multiple NQF-endorsed measures for which CDC reports measure results on behalf of healthcare facilities to the Centers for Medicare and Medicaid Services (CMS) quality measurement reporting programs.

- Colorado Department of Public Health and Environment Patient Safety Program - Healthcare-associated infections (HAI) are among the top ten leading causes of death in the United States. Colorado recognizes the seriousness of this public health problem and passed the HAI reporting legislation in 2006. House bill 1045 requires hospitals, hospital units, ambulatory surgery centers and dialysis centers to report healthcare-associated infections using the National Healthcare Safety Network (NHSN). This legislation created the Patient Safety Program at the Colorado Department of Public Health and Environment (CDPHE).
- This is a new measure. Its initial use for public health/disease surveillance, quality improvement with benchmarking (external benchmarking to multiple organizations), and quality improvement (internal to the specific organization) will enable the measure steward, the CDC's National Healthcare Safety Network (NHSN), to identify and address any gaps in the measure specifications that must be closed before CDC can recommend the measure for public reporting or other accountability purposes on the federal level.
- The CDC's National Healthcare Safety Network (NHSN) will work with ASCs that report SSI data to NHSN to further evaluate the measure's usefulness for SSI prevention and to refine the measure as needed to improve its value for assessing variation in SSI rates intra- and inter-organizationally. NHSN will serve as the data aggregating system. The NHSN Outpatient Procedure Component will provide the technical infrastructure for data collection, analysis, and measure results reporting to participating ASCs, including national benchmarks presented using the SIRs as the summary measures. This additional field experience with measure data, coupled with systematic studies, will serve to define what additional data and methods, if any, are needed to recommend use of this measure for accountability purposes on the federal level.

Improvement results:

Unexpected findings (positive or negative) during implementation:

- No unexpected findings reported.

Potential harms:

- No negative unintended consequences have been identified

Feedback :

- Developer did not identify any specific feedback loops related to this 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

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: ** Currently in use for quality reporting.-High
** This is a new measure. Its initial use for public health/disease surveillance, QI with benchmarking (internal and external), will enable the Surveillance Branch, Division of Healthcare Quality Promotion, CDC, to identify and address gaps in the measure specifications that must be closed before CDC recommends this for public reporting; the National Healthcare Safety Network (NHSN) and CDC are partners in clinical care and public health surveillance. This measure will provide actionable data for improving healthcare within and across ASCs.
** Should be publicly reported. Good usability potential, both for facility use and public decision-making.
** The measure is already being publicly reported by Colorado Department of Public Health and Environment Patient Safety Program through NHSN.
** The measure is publicly reported and used in accountability program. (NHSN)

Criterion 5: Related and Competing Measures
Related or competing measures N/A Harmonization N/A

Pre-meeting public and member comments

-

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Title: Ambulatory Breast Procedure Surgical Site Infection (SSI) Outcome Measure

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

Date of Submission:

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

Subcriterion 1a. Evidence to Support the Measure Focus

The measure focus is a health outcome or is evidence-based, demonstrated as follows:

- Health outcome:[3] a rationale supports the relationship of the health outcome to processes or structures of care.
- Intermediate clinical outcome, Process,[4] or Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence[5] that the measure focus leads to a desired health outcome.
- Patient experience with care: evidence that the measured aspects of care are those valued by patients and for which the patient is the best and/or only source of information OR that patient experience with care is correlated with desired outcomes.
- Efficiency:[6] evidence for the quality component as noted above.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.
4. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement.
5. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.
6. Measures of efficiency combine the concepts of resource use and quality (NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1. This is a measure of:

Outcome

☒ Health outcome: Click here to name the health outcome

Health outcome includes patient-reported outcomes (PRO, i.e., HRQoL/functional status, symptom/burden, experience with care, health-related behaviors)

☐ Intermediate clinical outcome: Click here to name the intermediate outcome

☐ Process:

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME PERFORMANCE MEASURE *If not a health outcome, skip to 1a.3*

1a.2. Briefly state or diagram the linkage between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

See literature review in answers below.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) and at least one healthcare structure, process, intervention, or service.

See literature review in answers below.

Note: For health outcome performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the linkages between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation

Draft Guideline—Centers for Disease Control and Prevention Draft Guideline for the Prevention of Surgical Site Infections January 2014: https://www.regulations.gov/#!documentDetail;D=CDC-2014-0003-0002

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

I. PARENTERAL ANTIMICROBIAL PROPHYLAXIS

1A. Administer preoperative antimicrobial agent only when indicated, based on published clinical practice guidelines and timed such that a bactericidal concentration of the agent is established in the serum and tissues when the incision is made **(Category IB)** 12 (Key Question 1A)
 No further refinement of timing can be made for preoperative antimicrobial agent based on clinical outcomes. **(No recommendation/unresolved issue)** (Key Question 1A)

1B. Administer the appropriate parenteral prophylactic antimicrobial agent prior to skin incision in all cesarean sections. **(Category IA)** 22-25 (Key Question 1B)
1C. No recommendation can be made regarding the safety and effectiveness of weight-adjusted dosing of parenteral prophylactic antimicrobial agents for the prevention of surgical site infection. **(No recommendation/unresolved issue)** (Key Question 1C)
1D. No recommendation can be made regarding the safety and effectiveness of intraoperative redosing of parenteral prophylactic antimicrobial agents for the prevention of surgical site infection. **(No recommendation/unresolved issue)** 26 (Key Question 1D) **DISCLAIMER:** This document is a DRAFT. The findings and conclusions in this draft guideline have not been formally disseminated by the Centers for Disease Control and Prevention and should not be construed to represent any agency determination or policy**.** 12

1E. In clean and clean-contaminated procedures, do not administer additional prophylactic antimicrobial agent doses after the surgical incision is closed in the operating room, even in the presence of a drain. **(Category IA)** 27-65 (Key Question 1E)

II. NON-PARENTERAL ANTIMICROBIAL PROPHYLAXIS

2A.1. No recommendation can be made regarding the safety and effectiveness of intraoperative antimicrobial irrigation (e.g., intra-abdominal, deep or subcutaneous tissues) for the prevention of surgical site infection. **(No recommendation/unresolved issue)** (Key Question 2A)

2A.2. No recommendation can be made regarding the safety and effectiveness of soaking prosthetic devices in antimicrobial solutions prior to implantation for the prevention of surgical site infection. **(No recommendation/ unresolved issue)** (Key question 2A)

2B.1. Do not apply antimicrobial agents (i.e., ointments, solutions, powders) to the surgical incision for the prevention of surgical site infection **(Category IB)** 66-72 (Key Question 2B)

2B.2. Application of autologous platelet rich plasma is not necessary for the prevention of surgical site infection. **(Category II)** 73-75 (Key Question 2B)

2C. Use of antimicrobial coated sutures is not necessary for the prevention of surgical site infection. **(Category II)** 76-79 (Key Question 2C)

2D. No recommendation can be made regarding the safety and effectiveness of antimicrobial dressings applied to surgical incisions following primary closure in the operating room for the prevention of surgical site infection. **(No recommendation/ unresolved issue)** (Key Question 2D)

III. GLYCEMIC CONTROL

3A.1. Implement perioperative glycemic control and use blood glucose target levels <200mg/dL in diabetic and non-diabetic patients. **(Category IA)** 80,81 (Key Question 3)

3A.2. No recommendation can be made regarding the safety and effectiveness of lower (<200mg/dL) or narrower blood glucose target levels, nor the optimal timing, duration, or delivery method of perioperative glycemic control for the prevention of surgical site infection. **(No recommendation/unresolved issue)** (Key Question 3)

3B. No recommendation can be made regarding optimal hemoglobin A1C target levels for the prevention of surgical site infection in diabetic and non-diabetic patients. **(No recommendation/unresolved issue)** (Key Question 3)

IV. NORMOTHERMIA

4. Maintain perioperative normothermia **(Category IA)** [82-84] (Key Question 4)

5. No recommendation can be made regarding the safety and effectiveness of strategies to achieve and maintain normothermia, the lower limit of normothermia, or the optimal timing and duration of normothermia for the prevention of surgical site infection. **(No recommendation/unresolved issue)** (Key Question 5)

V. OXYGENATION

6A. For patients with normal pulmonary function undergoing general anesthesia with endotracheal intubation, administer increased fraction of inspired oxygen (FiO_2) both intraoperatively and post-extubation in the immediate postoperative period. To optimize tissue oxygen delivery, maintain perioperative normothermia and adequate volume replacement. **(Category IA)** [85-90] (Key Question 6)

6B. No recommendation can be made regarding the safety and effectiveness of administering perioperative increased fraction of inspired oxygen (FiO_2) for the prevention of surgical site infection in patients with normal pulmonary function undergoing either general anesthesia without endotracheal intubation or neuraxial anesthesia (i.e., spinal, epidural, or local nerve blocks). **(No recommendation/unresolved issue)** [91] (Key Question 6)

6C. No recommendation can be made regarding the safety and effectiveness of administering increased fraction of inspired oxygen (FiO_2) via facemask or nasal cannula only during the postoperative period for the prevention of surgical site infection in patients with normal pulmonary function. **(No recommendation/unresolved issue)** [92,93] (Key Question 6)

7. No recommendation can be made regarding the optimal target level, duration, and delivery method of the fraction of inspired oxygen (FiO_2) for the prevention of surgical site infection. **(No recommendation/ unresolved issue)** (Key Question 7)

VI. ANTISEPTIC PROPHYLAXIS

8A. Advise patients to shower or bathe (full body) with either soap (antimicrobial or non-antimicrobial) or an antiseptic agent on at least the night before the operative day **(Category IB)** [94-102] (Key Question 8A)

8A.1. No recommendation can be made regarding the optimal timing of the preoperative shower or bath, the total number of soap or antiseptic agent applications, or the use of chlorhexidine gluconate washcloths for the prevention of surgical site infection. **(No recommendation/ unresolved issue)** (Key Question 8A)

8B. Perform intraoperative skin preparation with an alcohol-based antiseptic agent, unless contraindicated. **(Category IA)** 103-116 (Key Question 8B)
8C. Application of an antimicrobial sealant immediately following intraoperative skin preparation is not necessary for the prevention of surgical site infection. **(Category II)** 117-119 (Key Question 8C)
8D. Use of plastic adhesive drapes with or without antimicrobial properties, is not necessary for the prevention of surgical site infection. **(Category II)** 104,120-124 (Key Question 8D)
9A. Consider intraoperative irrigation of deep or subcutaneous tissues with aqueous iodophor solution for the prevention of surgical site infection. Intra-peritoneal lavage with aqueous iodophor solution in contaminated or dirty abdominal procedures is not necessary. **(Category II)** 125-131 (Key Question 9)
9B. No recommendation can be made regarding the safety and effectiveness of soaking prosthetic devices in antiseptic solutions prior to implantation for the prevention of surgical site infection. **(No recommendation/unresolved issue)** (Key Question 9)
10. No recommendation can be made regarding the safety and effectiveness of repeat application of antiseptic agents to the patient's skin immediately prior to closing the surgical incision for the prevention

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

*Category IA.*Strongly recommended for implementation and supported by well-designed experimental, clinical, or epidemiological studies.

*Category IB.*Strongly recommended for implementation and supported by some experimental, clinical, or epidemiological studies and strong theoretical rationale.

Category II. Suggested for implementation and supported by suggestive clinical or epidemiological studies or theoretical rationale.

No recommendation; unresolved issue. Practices for which insufficient evidence or no consensus regarding efficacy exists.

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☒ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1a. Deverick J. Anderson, Kelly Podgorny, Sandra I. Berríos-Torres, et al. Strategies to Prevent Surgical Site Infections in Acute Care Hospitals: 2014 Update. *Infection Control & Hospital Epidemiology*. 2014; 35: 605-627.

1a.6.1b. Nafziger, D. A., Lundstrom, T., Chandra, S., & Massanari, R. M. Infection control in ambulatory care. *Infectious Disease Clinics of North America*. 1997;11(2): 279-296.

1a.6.1c.

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

1a.6.1a (Deverick 2014)

Grade Definition

I. High: Highly confident that the true effect lies close to that of the estimated size and direction of the effect. Evidence is rated as high quality when there is a wide range of studies with no major limitations, there is little variation between studies, and the summary estimate has a narrow confidence interval.

II. Moderate: The true effect is likely to be close to the estimated size and direction of the effect, but there is a possibility that it is substantially different. Evidence is rated as moderate quality when there are only a few studies and some have limitations but not major flaws, there is some variation between studies, or the confidence interval of the summary estimate is wide.

III. Low: The true effect may be substantially different from the estimated size and direction of the effect. Evidence is rated as low quality when supporting studies have major flaws, there is important variation between

studies, the confidence interval of the summary estimate is very wide, or there are no rigorous studies, only expert consensus.

Note: Based on Grades of Recommendation, Assessment, Development, and Evaluation (GRADE) and the Canadian Task Force on Preventive Health Care.

Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ.* 2008;336(7650):924–926.

http://www.bmj.com/content/336/7650/924.full.pdf+html

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

1a.7.1a (Deverick 2014)

"9. Perform surveillance for SSI (quality of evidence: II).

a. Identify high-risk, high-volume operative procedures to be targeted for SSI surveillance on the basis of a risk assessment of patient populations, operative procedures performed, and available SSI surveillance data.

b. Identify, collect, store, and analyze data needed for the surveillance program.

i. Develop a database for storing, managing, and accessing data collected on SSIs.

ii. Implement a system for collecting data needed to identify SSIs. Data are required from surgical and microbiological databases. Obtain the following data from surgical databases: patient name, medical record number, date, type of procedure, surgeons, anesthesiologists, incision time, wound class, ASA score, closure time, and presence of an SSI. Ideally, these data are supplemented with process data, including prophylactic agent and dose and time(s) of administration of prophylactic agent. For patients diagnosed with an SSI, necessary microbiological data include type of SSI, infecting organism and antimicrobial susceptibilities, and date of infection. More detailed surgical and patient information may be useful for some procedures, including use of general anesthesia, emergency or trauma-related surgery, body mass index, and diagnosis of diabetes.

iii. Prepare periodic SSI reports (time frame will depend on hospital needs and volume of targeted procedures).

iv. Collect denominator data on all patients undergoing targeted procedures in order to calculate SSI rates for each type of procedure.

v. Identify trends (eg, in SSI rates and pathogens causing SSIs).

c. Use updated CDC NHSN definitions for SSI.

d. Perform indirect surveillance for targeted procedures.

e. Perform postoperative surveillance for 30 days; extend the postoperative surveillance period to 90 days for certain procedure categories.

i. Procedures that require 90-day surveillance are determined by specific procedure codes.

f. Surveillance should be performed on patients readmitted to the hospital.

i. If an SSI is diagnosed at your institution but the surgical procedure was performed elsewhere, notify the hospital where the original procedure was performed.

g. Develop a system for routine review and interpretation of SSI rates to detect significant increases or outbreaks and to identify areas where additional resources might be needed to improve SSI rates. If increased rates are identified, determine the number of potentially preventable infections that occurred, defined as the number of SSIs that occurred during a procedure in which less than 100% of recommended practices and processes were completed." (Page 611)

"11. Provide ongoing feedback of SSI rates to surgical and perioperative personnel and leadership (quality of evidence: II).

a. Routinely audit and provide confidential feedback on SSI rates and adherence to process measures to individual surgeons, the surgical division and/or department chiefs, and hospital leadership.17,130

i. For each type of procedure performed, provide risk-adjusted rates of SSI.

ii. Anonymously benchmark procedure-specific risk-adjusted rates of SSI among peer surgeons." (Page 611-12)

1a.7.1b (Nafziger 1997)

Main Findings: This review aggregated 12 surveillance studies and observed rates ranging from 0 to 16 infections per 100 procedures. Surveillance methods were consistently poor and the variability between estimates was high. From the evidence summarized in the review, the authors concluded that information regarding the risk for SSI in ASCs and the cost effectiveness of alternative surveillance strategies was still too inconclusive to make valid inferences.

1a.7.2. Grade assigned for the quality of the quoted evidence <u>with definition</u> of the grade:

See above

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).

Date range: no date limit-1999 (Mangram 1999)

Date range: no date limit-2014 (Deverick 2014)

Date range: no date limit-1997 (Nafziger 1997)

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

1a.7.5a (Mangram 1999)

Guidelines recommend surveillance- number and types of studies used to inform guideline not specified.

1a.7.5b (Deverick 2014)

Guidelines recommend surveillance- number and types of studies used to inform guideline not specified.

1a.7.5c(Nafziger 1997)

7 papers, 12 surveillance studies stratified by surveillance method type and patient population (determined by wound classification). Surveillance methods included Physician/patient mail surveys, patient phone surveys, computer-based physician survey, clinical examination at follow up, or a combination of methods.

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

The overall body of evidence on the incidence, outcomes, and prevention of SSIs in the ASC patient population is sparse but the available data suggest risks for SSIs following some breast procedures in some settings may be as high as 30%.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

In the current literature, the rates of SSI in ambulatory surgery centers is relatively low—however, aggregate numbers of infections can still cause a substantial burden, as those often result in post-surgical visits and morbidity. ASCs have been shown to have a lower SSI rate than inpatient settings; in one study, SSI morbidity

and recurrence rates in ambulatory surgery were half the rates in inpatient surgery. A 5-year study of SSIs in ambulatory surgery centers showed a rate of 2.8 SSI per 100 surgeries (Vilar-Compte 2001). These rates are relatively consistent- another study reported a risk of SSI after outpatient surgery to be 3.5% (Grøgaard 2001). Aside from morbidity alone, postsurgical visits due to SSI acquired during surgery contribute much to the cost burden on healthcare facilities. A study on postsurgical acute care visits for SSIs in ASCs demonstrated a rate of 3.09 SSI-related visits per 1000 procedures at 14 days after surgery and 4.84 per 1000 at 30 days after surgery (Owens 2014).

Though estimates of risk for breast procedures specifically vary from 1% to over 30% (and rate varies from 3 SSI to 28 SSI per 1000 procedures) depending on breast procedure type, sample population, and definition of SSI, it is clear that breast procedure-related SSIs are a large burden to outpatient healthcare facilities, and provide much room for benefit. There is little data on the number or proportion of preventable SSI specifically following breast procedures conducted in ASCs.

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

No harms of surveillance and measurement of breast procedure-related SSI in ASCs were mentioned in any of the systematic reviews.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
Ambulatory_Breast_Procedure_SSI_Outcome_Measure_Proposal_Evidence_Attachment_05.31.2016.docx

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., the benefits or improvements in quality envisioned by use of this measure)*
The measure provides summary results that ASCs can use as quantitative aids in their efforts to evaluate and reduce breast surgery surgical site infection rates. The SIRs can be used by ASCs to benchmark SSI rates, identify opportunities for improvement, and gauge the impact of prevention efforts. At the outset, the SIRs provide a set of signals that often warrant further analysis, such as an examination of lapses in infection control practices that may contribute to high incidence of SSI. Some of the analytic follow up can be completed with data reported to CDC's National Healthcare Safety Network (NHSN)Patient Safety Component Procedure-Associated (PA) Module, using analytic features built into the NHSN application. However, additional analyses to determine the cause of infections as targets for prevention in individual instances are likely to require access to data that is beyond the scope of data collection and analysis using the NHSN module.

Breast procedures were specifically chosen for this measure due to the observed burden of breast procedure-associated SSI. Out of 67,150 ASC procedures reported to NHSN from 2010-2013, 30,787 (45.9%) were breast procedures. Out of the 142 SSIs reported from ASCs during the same time period, 78 (54.9%) were related to breast procedures, indicating an SSI risk of 0.25%. This was the highest volume and SSI risk out of all outpatient ASC procedures reported in the timeframe.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
An exploratory analysis of NHSN data showed that out of 67,150 ASC procedures reported to NHSN from 2010-2013, 30,787 (45.9%) were breast procedures. Out of the 142 SSIs reported from ASCs during the same time period, 78 (54.9%) were related to breast procedures, indicating a risk of SSI of 0.25%. This was the highest volume and SSI risk among all outpatient ASC procedures reported in the timeframe.

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.
Numerous individual studies and systematic reviews provide strong evidence that measurement and feedback of surgical site infections leads to lower SSI rates in the long term. (Anderson 2014, Mangram 1999, Gaynes 2001, Vilar-Compte 2009). Although standardized metrics have been developed to measure SSI rates for inpatient surgeries in the hospital setting (Mu 2009), these have not yet been developed for outpatient surgeries in ASCs, which comprise a fast-growing proportion of all surgeries performed in the US (Kozak 1999). The measure will serve as a quantitative guide for ASCs, enabling them to benchmark SSI rates in their facilities against nationally aggregated data and set targets for improvement.

Citations:

Anderson, Deverick J, Kelly Podgorny, Sandra I. Berríos-Torres, et al. Strategies to Prevent Surgical Site Infections in Acute Care Hospitals: 2014 Update. Infection Control & Hospital Epidemiology. 2014; 35: 605-627.

Mangram AJ, Horan TC, Pearson ML, Silver LC, Jarvis WR. Guideline for prevention of surgical site infection, 1999. Hospital Infection Control Practices Advisory Committee. Infect Control Hosp Epidemiol. 1999;?20: 250-278.
http://www.cdc.gov/hicpac/pdf/guidelines/SSI_1999.pdf.
Gaynes R, Richards C, Edwards JR, et al. Feeding back surveillance data to prevent hospital-acquired infections. Emerg Infect Dis. 2001; 7: 295–298.
Vilar-Compte, D., Rosales, S., Hernandez-Mello, N., Maafs, E., & Volkow, P. Surveillance, control, and prevention of surgical site infections in breast cancer surgery: a 5-year experience. American journal of infection control. 2009; 37(8): 674-679.
Mu, Y., et al. Improving risk-adjusted measures of surgical site infection for the national
healthcare safety network. Infect Control Hosp Epidemiol. 2011; 32(10): 970-86.

Kozak LJ, McCarthy E, Pokras R. Changing patterns of surgical care in the United States, 1980-1995. Health Care Financ Rev. 1999; 21(1): 31-49.

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*

Age and Gender Disparities in Surgical Site Infections (SSIs) among Outpatient Surgical Breast Procedures, Reported to NHSN, 2010-2013

Variable	No. Procedures	No. SSIs	Risk (%)	P (Likelihood Ratio)
Age Quartiles				< 0.0001
< 40 years	8071	3	0.04	
41-51 years	7546	16	0.21	
52-62 years	7875	32	0.41	
> 62 years	7175	27	0.38	
Gender				0.0414
Female	30001	78	0.26	
Male	810	0	0	

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**

Data on disparities in surgical site infections in ASCs, as well as in hospitals, are sparse. No studies or reviews were found specifically on disparities surrounding SSI in any healthcare facility. However, it has been extensively documented that surgical site infections lead to an excess cost burden as well as excess hospital stay for patients (Zimlichman 2013, Olsen 2008, Kirkland 1999). These additional costs may cause disparities in care for SSI, which are reflective of disparities in access to health care in general (Brown 2000, Lasser 2008).

Citations:

Zimlichman E, Henderson D, Tamir O, et al. Health Care–Associated Infections: A Meta-analysis of Costs and Financial Impact on the US Health Care System. JAMA Intern Med. 2013;173(22):2039-2046.

Olsen MA, et al. Hospital-Associated Costs Due to Surgical Site Infection After Breast Surgery. Arch Surg. 2008; 143(1): 53-60.

Kirkland KB, Briggs JP, Trivette SL, Wilkinson WE, Sexton DJ. The impact of surgical-site infections in the 1990s: attributable mortality, excess length of hospitalization, and extra costs. Infect Control Hosp Epidemiol. 1999;?20:725-730.

Brown, E. R., Ojeda, V. D., Wyn, R., & Levan, R. (2000). Racial and ethnic disparities in access to health insurance and health care. UCLA Center for Health Policy Research.

Lasser, Karen E., David U. Himmelstein, and Steffie Woolhandler. "Access to care, health status, and health disparities in the United States and Canada: results of a cross-national population-based survey." Health Policy: Crisis and Reform in the US Health Care Delivery System (2008): 379.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**

Affects large numbers, Frequently performed procedure, High resource use, Patient/societal consequences of poor quality, Severity of illness

1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**

Healthcare-associated infections (HAIs) are a major cause of morbidity and mortality in healthcare settings in the United States. The most recent prevalence surveys of HAIs have estimated that approximately 4.0% of inpatients in acute care settings have developed at least one HAI, translating to 721,800 infections in 648,000 patients in 2011 (Magill 2014). One meta-analysis including studies from 1986-2013 reported an annual HAI cost of $9.8 billion. SSIs contributed to most of the total cost (33.7%), followed by VAP (31.6%), CLABSI (18.9%), C. difficile infections (18.4%), and CAUTI (<1%) (Zimlichman 2013).

Surgical site infection is one of the most common healthcare-associated infections, comprising approximately 22% of all HAIs (Magill 2014). In addition to being a highly prevalent type of HAI, SSIs also contribute greatly to the mortality and cost burden of HAIs. A 1999 study of 255 pairs of patients with and without SSI matched on age, procedure, NNIS risk index, date of surgery, and surgeon found that those with SSIs have twice the mortality rate of those without SSI and are five times as likely to be readmitted to the hospital. The mean excess hospital stay directly attributable to SSI was 12 days, and the excess costs attributable to SSI were approximately $5,000 per patient (Kirkland 1999). More recent studies have estimated the costs of an SSI to be even higher—a 2007 study cited a range from approximately $11,000 to $35,000 per SSI (Scott 2009), and one meta-analysis of healthcare costs in 2013 determined the cost to be $20,785 per SSI (Zimlichman 2013).

Breast SSIs contribute a substantial portion of SSI in inpatient settings, and also have the one of the highest risk of any procedure type in outpatient settings. In the Netherlands, the rate of SSI following mastectomies in 2006 was 61% as determined by a study in 2006 (Mannien 2006). A case control study performed in 2004 reported SSI rates following breast surgeries to be 25.8% (Vilar-Compte 2004). One study of breast SSI risk in an HOPD reported an overall risk of 5.2%, with procedure-specific risks of 12.4% following mastectomy with immediate implant reconstruction, 6.2% following mastectomy with immediate reconstruction using a transverse rectus abdominis myocutaneous flap, 4.4% following mastectomy only, and 1.1% following breast reduction surgery (Olsen 2008). Another study of SSI following breast cancer-related procedures reported a risk of 18.9% (Vilar-Compte 2009). The cost incurred by each breast SSI attributable to the SSI was estimated by one analysis to be $4,901 per patient (Olsen 2008). Though these estimates of risk vary from 1% to over 30% depending on procedure type, sample population, and definition of SSI, it is clear that breast procedure-related SSIs are a large burden to outpatient healthcare facilities.

From 1980-1995, a significant trend in surgery was the transition from inpatient settings to outpatient ambulatory surgery settings due to advances in surgical techniques and economic incentives for ambulatory surgery (Kozak 1999). In the current literature, the rates of SSI in ambulatory surgery centers is relatively low—however, aggregate numbers of infections can still cause a substantial burden, as those often result in post-surgical visits and morbidity. ASCs have been shown to have a lower SSI rate than inpatient settings; in one study, SSI morbidity and recurrence rates in ambulatory surgery were half the rates in inpatient surgery. A 5-year study of SSIs in ambulatory surgery centers showed a rate of 2.8 SSI per 100 surgeries (Vilar-Compte 2001). These rates are relatively consistent- another study reported a risk of SSI after outpatient surgery to be 3.5% (Grøgaard 2001). Aside from morbidity alone, postsurgical visits due to SSI acquired during surgery contribute much to the cost burden on healthcare facilities. A study on postsurgical acute care visits for SSIs in ASCs demonstrated a rate of 3.09 SSI-related visits per 1000 procedures at 14 days after surgery and 4.84 per 1000 at 30 days after surgery (Owens 2014).

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

Magill SS, Edwards JR, Bamberg W, Beldavs ZG, Dumyati G, Kainer MA, Multistate point-prevalence survey of health care–associated infections. N Engl J Med. 2014; 370: 1198–208.

Zimlichman E, Henderson D, Tamir O, et al. Health Care–Associated Infections: A Meta-analysis of Costs and Financial Impact on the US Health Care System. JAMA Intern Med. 2013;173(22):2039-2046.

Kirkland KB, Briggs JP, Trivette SL, Wilkinson WE, Sexton DJ. The impact of surgical-site infections in the 1990s: attributable mortality, excess length of hospitalization, and extra costs. Infect Control Hosp Epidemiol. 1999;?20:725-730.

Scott II RD. The direct medical costs of healthcare-associated infections in U.S. hospitals and the benefits of prevention. 2009. http://www.cdc.gov/ncidod/dhqp/pdf/Scott_CostPaper.pdf. Accessed August 12, 2011.

Mannien J, Wille JC, Snoeren RL, van den Hof S. Impact of postdischarge surveillance on surgical site infection rates for several surgical procedures: results from the nosocomial surveillance network in The Netherlands. Infect Control Hosp Epidemiol. 2006;?27: 809-816.

Vilar-Compte, D., Jacquemin, B., Robles-Vidal, C., & Volkow, P. (2004). Surgical site infections in breast surgery: case-control study. World journal of surgery. 2004; 28(3): 242-246.

Vilar-Compte, D., Rosales, S., Hernandez-Mello, N., Maafs, E., & Volkow, P. Surveillance, control, and prevention of surgical site infections in breast cancer surgery: a 5-year experience. American journal of infection control. 2009; 37(8): 674-679.

Olsen MA, et al. Hospital-Associated Costs Due to Surgical Site Infection After Breast Surgery. Arch Surg. 2008; 143(1): 53-60.

Kozak LJ, McCarthy E, Pokras R. Changing patterns of surgical care in the United States, 1980-1995. Health Care Financ Rev. 1999; 21(1): 31-49.

Vilar-Compte D, Roldán R, Sandoval S, et al. Surgical site infections in ambulatory surgery: a 5-year experience. Am J Infect Control. 2001; 29(2): 99-103.

Grøgaard B, Kimsås E, Raeder J. Wound infection in day-surgery. Ambul Surg. 2001; 9(2): 109-112.

Owens PL, Barrett ML, Raetzman S, Maggard-Gibbons M, Steiner CA. Surgical Site Infections Following Ambulatory Surgery Procedures. JAMA. 2014; 311(7): 709-716.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
Does not apply

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 subcriteria 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):*
Infectious Diseases, Surgery

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Prevention, Safety : Healthcare Associated Infections

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.)*
NHSN webpage with specific information to be provided as part of the forthcoming NHSN Outpatient Proceudre Component

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

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:**
Breast_Procedure_CPT_List_and_Final_Model_for_Ambulatory_Breast_Procedure_SSI_Outciome_Measure_05.31.2016_-

_Copy.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date 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)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Surgical site infections (SSIs) during the 30-day (superficial SSI) and 90-day (deep and organ/space SSI) postoperative periods following breast procedures in Ambulatory Surgery Centers.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Data for this measure can be aggregated at multiple levels. For NHSN purposes, data will be aggregated quarterly and annually based on the calendar year. However, facilities or groups may choose to aggregate data at different intervals (monthly, fiscal year, etc.) for their own quality initiatives. This is made possible through the NHSN Analysis tool.

S.6. **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, 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.
SSIs are defined in the NHSN Patient Safety Protocol: http://www.cdc.gov/nhsn/CPTcodes/ssi-cpt.html.

Surgical site infection: An infection, following a breast procedure, of either the skin, subcutaneous tissue and breast parenchyma at the incision site (superficial incisional SSI), deep soft tissues of the incision site (deep incisional SSI), or any part of the body deeper than the fascial/muscle layers that is opened or manipulated during the operative procedure (organ/space SSI).

Superficial incisional SSI
Must meet the following criteria:
Infection occurs within 30 days after any NHSN operative procedure (where day 1 = the procedure date)
AND
involves only skin, subcutaneous tissue (e.g. fatty tissue) and breast parenchyma (e.g. milk ducts and glands that produce milk) of the incision
AND
patient has at least one of the following:
a. purulent drainage from the superficial incision.
b. organisms identified from an aseptically-obtained specimen
from the superficial incision or subcutaneous tissue by a culture or non-culture based microbiologic testing method which is performed for purposes of clinical diagnosis or treatment (e.g., not Active Surveillance Culture/Testing (ASC/AST).
c. superficial incision that is deliberately opened by a surgeon, attending physician** or other designee and culture or non-culture based testing is not performed.
d. diagnosis of a superficial incisional SSI by the surgeon or attending physician** or other designee.
AND
patient has at least one of the following signs or symptoms: pain or tenderness; localized swelling; erythema; or heat. A culture or non-culture based test that has a negative finding does not meet this criterion.

Deep incisional SSI
Must meet the following criteria:
Infection occurs within 90 days after the NHSN operative procedure (where day 1 = the procedure date)
according to the list in Table 2
AND
involves deep soft tissues of the incision (e.g., fascial and muscle layers)
AND

patient has at least one of the following:
a. purulent drainage from the deep incision.
b. a deep incision that spontaneously dehisces, or is deliberately opened or aspirated by a surgeon, attending physician** or other designee and organism is identified by a culture or non-culture based microbiologic testing method which is performed for purposes of clinical diagnosis or treatment (e.g., not Active Surveillance Culture/Testing (ASC/AST) or culture or non-culture based microbiologic testing method is not performed
c. an abscess or other evidence of infection involving the deep incision that is detected on gross anatomical or histopathologic exam, or imaging test
AND
patient has at least one of the following signs or symptoms: fever (>38°C); localized pain or tenderness. A culture or non-culture based test that has a negative finding does not meet this criterion.

Organ/Space SSI
Must meet the following criteria:
Infection occurs within 30 or 90 days after the NHSN operative procedure (where day 1 = the procedure date) according to the list in Table 2
AND
infection involves any part of the body deeper than the fascial/muscle layers (e.g. subpectoral), that is opened or manipulated during the operative procedure
AND
patient has at least one of the following:
a. purulent drainage from a drain that is placed into the organ/space (e.g., closed suction drainage system, open drain, T-tube drain, CT guided drainage)
b. organisms are identified from an aseptically-obtained fluid or tissue in the organ/space by a culture or non-culture based microbiologic testing method which is performed for purposes of clinical diagnosis or treatment (e.g., not Active Surveillance Culture/Testing (ASC/AST).
c. an abscess or other evidence of infection involving the organ/space that is detected on gross anatomical or histopathologic exam, or imaging test
AND
meets at least one of the following criteria for BRST-Breast abscess or mastitis

BRST-Breast abscess/infection
1. Patient has organisms identified from affected breast tissue or fluid obtained by invasive procedure by a culture or non-culture based microbiologic testing method which is performed for purposes of clinical diagnosis or treatment (e.g., not Active Surveillance Culture/Testing (ASC/AST).
2. Patient has a breast abscess or other evidence of infection on gross anatomic or histopathologic exam.
AND
Physician initiates antimicrobial therapy within 2 days of onset or worsening of symptoms.

Notes:
- Breast procedures may involve a secondary operative site. i.e., procedures that include flaps. The flap site is the secondary site. Secondary sites have a 30 day surveillance period. If the secondary site meets criteria for an SSI, it reported as either a superficial incisional SSI at the secondary site or deep incisional infection at the incisional site.
- Accessing a breast expander after a breast procedure is considered an invasive procedure and any subsequent infection is not deemed an SSI attributable to the breast procedure.

** The term attending physician for the purposes of application of the NHSN SSI criteria may be interpreted to mean the surgeon(s), infectious disease, other physician on the case, emergency physician or physician's designee (nurse practitioner or physician's assistant).

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Breast procedures, as specified by the operative codes that comprise the breast procedure category of the NHSN Patient Safety Component Protocol, performed at ambulatory surgery centers.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
Information required to calculate the denominator:

CPT codes for NHSN Breast Procedure category:
11970, 19101, 19112, 19120, 19125, 19126, 19300, 19301, 19302, 19303, 19304, 19305, 19306, 19307, 19316, 19318, 19324, 19325, 19328, 19330, 19340, 19342, 19350, 19355, 19357, 19361, 19364, 19366, 19367, 19368, 19369, 19370, 19371, 19380

See attached spreadsheet for descriptions of each code.

Note: Bilateral breast procedures performed during the same trip to operating room are counted as two separate procedures

Ambulatory surgical center (ASC): any distinct entity that operates exclusively for the purpose of providing surgical services to patients not requiring hospitalization and in which the expected duration of services would not exceed 24 hours following an admission.

Parameter estimates for breast procedure logistic regression model are needed to calculate the expected number of SSIs (included in the attached document).

Patient-specific data: Age, American Society of Anesthesiologists Physical Status Classification (ASA Class).

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Hospital inpatients and hospital outpatient department patients, pediatric patients and very elderly patients, and brain-dead patients whose organs are being removed for donor purposes

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
Exclusion Criteria:
1. Inpatient breast procedures*
2. Breast procedures performed on patients under age 18 or age 109 or over.
3. Breast procedures with ASA Class VI (6).

*Breast procedures performed in hospital outpatient departments (HOPDs) are not included in the measure scope.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
None

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Multivariable logistic regression modeling including factors associated with differences in risk of surgical site infection. Variables available and considered in modeling: Patient age, ASA class, duration of procedure, Patient gender, wound classification, anesthesia use. Final risk model: Patient Age, ASA class.

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

Available in attached Excel or csv file at S.2b

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)

S.16. **Type of score:**
Ratio
If other:

S.17. **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.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
Each SIR is calculated as follows:
1. Identify the number of infections reported during the measurement period for an observed number of infections.
2. Obtain the predicted number of infections by applying the risk adjustment model to all eligible breast procedures during the measurement period.
3. Divide the observed number of infections by the predicted number of infections.
4. Result = SIR for the given period.
5. Note: SIRs are not calculated when the number of predicted infections is less than 0.2.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Does not apply

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Does not apply

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Procedures and SSIs with missing data were not included in the analysis

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Paper Medical Records

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Currently, NHSN data collection for SSIs following outpatient operative procedures is via the Patient Safety Component. Plans call for NHSN data collection for SSIs following outpatient operative procedures to be moved to the new Outpatient Procedure Component in 2018.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Ambulatory Surgery Center (ASC)
If other:

S.28. **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.)*

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
Ambulatory_Breast_Procedure_SSI_Outcome_Measure_Measure_Testing_Attachment_05.31.2016-636002891868282125.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Ambulatory Breast Procedure Surgical Site Infection (SSI) Outcome Measure
Date of Submission: Click here to enter a date
Type of Measure:

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

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed

performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score

as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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.23***)	**Measure Tested with Data From:**
☒ abstracted from paper record	☒ abstracted from paper record
☐ administrative claims	☐ administrative claims
☒ clinical database/registry	☒ clinical database/registry
☒ abstracted from electronic health record	☒ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other:	☐ other: Click here to describe

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

Data for model validation and reliability testing was obtained from breast surgical procedures conducted at ambulatory surgery centers (ASCs) reported to CDC's National Healthcare Safety Network (NHSN). Model validation used the entire dataset (January 2010-December 2014), and reliability testing focused on procedures reported from selected ASCs in Colorado from January to December 2014.

1.3. What are the dates of the data used in testing? January 2010 – December 2014

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.26*)	Measure Tested at Level of:
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

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

Risk model creation/testing dataset:

Table 1. Descriptive Characteristics of Ambulatory Surgery Centers reporting Outpatient Surgical Breast Procedures to NHSN, 2010-2014 (n=138).

Variable	n(%) or Med (Q1,Q3)
State	
New Jersey	70 (50.7%)
Colorado	45 (32.6%)
New Hampshire	10 (7.2%)
Texas	6 (4.3%)
Wisconsin	3 (2.2%)
Nevada	2 (1.4%)
Missouri	1 (0.7%)
Mississippi	1 (0.7%)
Procedure Volume	112 (16, 417)

Reliability testing facility characteristics:

Table 2: Distribution of Facilities by City

City	# Facilities
Denver Metro	7
Colorado Springs	5

Fort Collins	2
Loveland	2
Pueblo	1
Boulder	1

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

Risk Model validation dataset characteristics:

Table 3. Descriptive Characteristics of Predictors of Surgical Site Infections among Outpatient Surgical Breast Procedures Performed in Ambulatory Surgery Centers Reported to NHSN, 2010-2014 (n=37,673).

Variable	n(%) or Mean (SD)
Surgical Site Infections	93 (0.25%)
Age of Patient	47.7 (15.2)
Duration of Procedure	62.3 (48.0)
ASA Classification	
1	14,518 (38.5%)
2	19,536 (51.9%)
3/4/5	3619 (9.6%)
Wound Classification	
Clean/Clean Contaminated	37,500 (99.5%)
Contaminated/Dirty	172 (0.46%)
Anesthesia Used	30,260 (80.3%)
Female Gender	26,811 (97.7%)

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.

Risk-adjustment/model validation/exclusions: Sample used was breast procedures conducted in ASCs reported to NHSN from January 2010 - December 2014.

Reliability Testing: Sample used was breast procedures from Colorado ASCs, January-December 2014.

Validity Testing: The Ambulatory Surgery Center Quality Collaboration (ASC QC) administered a questionnaire, which included questions related to the four measure attributes, to 11 individuals currently working in ambulatory surgery centers (ASC). Seven respondents were registered nurses, working in regional operations, administration, clinical management, information technology, or quality improvement. The remaining 4 respondents were medical doctors; 2 general surgeons and 2 plastic surgeons.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

None.

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*)

Staff from the Colorado Department of Public Health and Environment conducted a reliability testing study of breast surgeries performed at ambulatory surgical centers (ASC) during calendar year 2014. From August through September 2015, staff conducted chart reviews for patients having breast surgeries in 18 Colorado ASCs. Selected ASC had performed at least 100 breast surgeries in 2014 and were located in the Denver metro area and along the Front Range (within 100 miles of Denver; Table 1). A total of 715 charts were examined (701 female and 14 male) to identify under- and over-reported events and data discrepancies and omissions in events and procedures. Table 2 presents an overview of NHSN reported procedures and events, number of charts reviewed, ineligible procedures, and under- and over-reported events for each ASC.

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*)

FINDINGS:

Under- and over-reported events. No under-reported events were found and one over-reported event was identified because the case did not meet all NHSN criteria for superficial SSI.

Data Discrepancies and Omissions

Bilateral procedures. Five facilities failed to enter two denominator forms in NHSN for bilateral procedures. These five entered the total procedure duration for a single bilateral procedure. The facilities were instructed to enter two denominator forms with exact durations for each procedure or the total time for both split equally between the two.

Procedure duration. As shown in table 4, 17 facilities (95%) reported procedure duration incorrectly in NHSN, because they used the protocol definition prior to the 2014 change. This variable was not included in the final model for reasons explained in 2b4.3.

ASA score. All 18 facilities had ASA score information available on their medical records and had reported the ASA variable into NHSN. Most ASA score discrepancies resulted from data entry errors.

Wound classification. Two facilities did not have wound classification information available in their medical records, but since it is a required variable, they reported the variable as "clean" for every case. This variable was not included in the final model for reasons explained in 2b4.3.

Table 4: Facility overview

ORGID	Breast procedures in 2014	# Charts Reviewed	# Reported SSI	# Charts with ineligible Procs	# Under-Reported Events	# Over-Reported Events
13437	304	40		17	0	0
13551	306	40	1	3	0	0
13692	513	40		0	0	0
13703	426	40		1	0	0
13730	175	39	1	2	0	1
13801	376	41	1	0	0	0
13988	180	40		3	0	0
14130	212	40		0	0	0
14153	174	40		2	0	0
14217	212	40		1	0	0
14542	148	40		9	0	0
14903	119	36		6	0	0
20974	229	40	1	0	0	0
21007	105	40		3	0	0
21040	222	41	1	0	0	0
29839	222	40		1	0	0
30467	649	39		5	0	0
34052	270	39		0	0	0
Total	**4842**	**715**	**5**	**53**	**0**	**1**

Table 5: Discrepant variables by facility

	Discrepant Variables				
ORGID	General Anesthesia	DOB	Wound Class	ASA Score	Procedure Time
13437	0	0	0	8	40
13551	1	0	0	11	40
13692	11	0	0	0	40
13703	0	0	1	0	40
13730	0	1	2	4	39
13801	0	0	2	0	41
13988	0	0	0	0	6

14130	0	0	0	0	40
14153	0	1	0	2	40
14217	0	0	0	10	40
14542	0	0	0	0	40
14903	6	0	30	6	36
20974	1	0	0	3	40
21007	0	0	37	0	40
21040	0	0	0	0	41
29839	2	1	2	4	40
30467	0	1	0	3	39
34052	0	0	0	0	39
Total (% discrepant)	21 (3%)	4 (<1%)	74 (10%)	51 (7%)	681 (95%)

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 measure is highly reliable, in being able to consistently and correctly identify SSI across facilities and raters. The number of data entry errors in ASA Score is relatively low (7%), and the number of errors in date of birth (Age) is very low (<1%), indicating that a performance score calculated using these elements would be reliable as well.

2b2. VALIDITY TESTING

2b2.1. 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 <u>performance measure score</u> 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*)

2b2.2. 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)*

The Centers for Disease Control and Prevention (CDC), working with the Ambulatory Surgery Center Quality Collaboration (ASC QC), developed a measure to assess the incidence of surgical site infections (SSI) following breast procedures. The validity, feasibility, interpretability, and actionability of the measure were assessed through a formal consensus process. Specifically, ASC QC administered a questionnaire, which included questions related to the four measure attributes, to 11 individuals currently working in ambulatory surgery centers (ASC). Seven respondents were registered nurses, working in regional operations, administration, clinical management, information technology, or quality improvement. The remaining 4 respondents were medical doctors; 2 general surgeons and 2 plastic surgeons. The questionnaire rated the respondent's level of agreement with statements related to each measure attribute based on a 5-point Likert Scale with a rating of 5 expressing agreement and 1 expressing disagreement. It also allowed respondents to elaborate on their ratings in open-ended questions. All respondents provided complete numeric ratings of the measure characteristics, and several respondents provided comments in open-ended questions. Questionnaire responses were analyzed to

assess the panel's consensus with respect to the validity, feasibility, interpretability, and actionability of the measure.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Validity Testing

There was high level of agreement among the respondents regarding the validity of the measure. Out of 11 respondents, 9 (81.8%) agreed that the measure appears to measure what it is intended to, giving a 5/5 rating response. The other two respondents rated their level of agreement with this statement with a 4/5 rating. Regarding the statement on whether the measure allows for consistent interpretation across centers, 9 out of 11 (81.8%) respondents agreed with a 5/5 rating, and 1 provided a 4/5 rating. The remaining respondent gave a 3/5 rating and expressed the difficulty inherent in dividing breast surgery SSI into categories of superficial and deep incisional due to the nature of the procedure.

The questionnaire also inquired about the extent to which the measure's score accurately reflects the quality of a center's performance. The majority of respondents (8 out of 11) agreed with the statement with a rating of 4/5 or 5/5; 2 neither agreed nor disagreed (3/5); and 1 disagreed (1/5). Several respondents elaborated that factors other than the quality of a center's performance, such as patient comorbidities, risk factors, and the quality of a surgeon can influence SSI. Regarding the statement that the measure's score can be used to distinguish between good and poor performance, 7 respondents (63.6%) agreed, giving a minimum rating of 4/5, 3 (27.3%) gave a rating of 3/5, and 1 disagreed with the statement (1/5). Several respondents again noted that SSI cannot be solely attributed to the quality of a center, and factors outside a facility's control, such as patient comorbidities, poor hygiene, and non-compliance with post-op instructions, may affect the measure's score.

The statements related to validity are listed below. Each statement was measured on a 5-point Likert Scale with a rating of 5 expressing agreement and a rating of 1 expressing disagreement.

- *The measure appears to measure what it is intended to.* (Median: 5.0/5.0; Mean: 4.8/5.0)
- *The measure is defined in a way that will allow for consistent interpretation of the inclusion and exclusion criteria from center to center.* (Median: 5.0/5.0; Mean: 4.7/5.0)
- *The measure score is an accurate reflection of the quality of center performance.* (Median: 4.0/5.0; Mean: 3.6/5.0)
- *The measure score can be used to distinguish good from poor performance.* (Median: 4.0/5.0; Mean: 3.6/5.0)

Feasibility Testing

In addition to validity, the questionnaire inquired about the feasibility of the measure with respect to effort and cost. The majority of respondents expressed agreement that data for the measure could be obtained with reasonable effort (81.8% with a minimum rating of 4/5) and reasonable cost (90.9% with a minimum rating of 4/5). In their open-ended responses, respondents noted the need for more patient engagement and increased labor costs to obtain the required data. One respondent indicated that 90 days is a difficult measure in ASCs.

The statements related to feasibility are listed below. Each statement was measured on a 5-point Likert Scale with a rating of 5 expressing agreement and a rating of 1 expressing disagreement.

- *The data required for the measure are likely to be obtained with reasonable effort.* (Median: 4.0/5.0; Mean: 4.2/5.0)
- *The data required for the measure are likely to be obtained with reasonable cost.* (Median: 4.0/5.0; Mean: 4.4/5.0)

Interpretability and Actionability Testing

All respondents agreed that providers can understand the results of the measure, giving a minimum rating of 4/5 to the relevant statement. The questionnaire responses also indicated that the measure is actionable. The majority of respondents (10 out of 11) agreed that a provider can take action based on measure results, with 8 respondents giving a 5/5 rating and 2 giving a 4/5 rating. One respondent gave a 2/5 rating. Regarding the existence of a direct linkage between the measure and improving the outcome/processes of care, 10 out of 11 respondents agreed with at least a 4/5 rating while 1 respondent gave a 2/5 rating. In response to the associated open-ended question, one of the respondents indicated some apprehension in the implementation of measures related to SSI due to the role of patient compliance in the prevention of SSI.

The statements related to interpretability and actionability are listed below. Each statement was measured on a 5-point Likert Scale with a rating of 5 expressing agreement and a rating of 1 expressing disagreement.

- *A provider can understand the results of the measure.* (Median: 5.0/5.0; Mean: 4.8/5.0)
- *If necessary, a provider can use the results of the measure to take action.* (Median: 5.0/5.0; Mean: 4.6/5.0)
- *This measure has a direct link to improving the outcome and/or related processes of care.* (Median: 5.0/5.0; Mean: 4.4/5.0)

2b2.4. 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?*)

See 2b2.3.

There was high level of agreement among the respondents regarding the validity of the measure. In addition to validity, the majority of respondents expressed agreement that data for the measure could be obtained with reasonable effort and reasonable cost. All respondents agreed that providers can understand the results of the measure.

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

2b3.1. 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*)

Procedures were excluded if age of patient was under 18 (pediatric population) or greater than 109 years (outlier values.

2b3.2. 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*)

No statistical testing performed on exclusions.

2b3.3. 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 population was limited to adult patients, given the nature of breast surgeries. Ages entered above 109 years were considered data entry errors and excluded from the population. All exclusions were necessary to achieve the most accurate and applicable model.

2b4. 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 2b5.

2b4.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**
☒ **Statistical risk model with 2 risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use 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.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

1. Potential adjustment factors were limited by the scope of variables collected by NHSN. Those considered, based on factors identified in literature, were: age of patient, anesthesia use, ASA classification, duration of procedure, gender of patient, and surgical wound classification (see Table 1).
2. Univariate analyses were conducted between each of these factors and the outcome to determine if the association was significant. Statistically significant univariate associations led to inclusion in the modeling process (all were significant).
3. Modeling process involved a backwards elimination of predictors from the saturated model. In each iteration, the least significant predictor was removed from the model until all remaining factors were significant. Other subsets of predictors were also considered. Duration of procedure was a significant factor, but was excluded because of clinical concerns about eligibility as a confounding factor.
4. The final model adjusted for categorical ASA classification, and ordinal age categories.

2b4.4a. What were the statistical results of the analyses used to select risk factors?

Final risk adjustment model:

Table 6. Final Model to Predict Surgical Site Infections (n=93) among Outpatient Surgical Breast Procedures Performed in Ambulatory Surgery Centers (n=37,673) Reported to NHSN, 2010-2014.

Effect	Estimate (SE)	Odds	95% CI	c-	H-L P

		Ratio		Index	
Intercept	-7.5840 (0.36)			0.675	0.6626
ASA Score					
2 vs. 1	0.9862 (0.33)	2.68	1.40, 5.12		
3/4/5 vs. 1	1.6882 (0.38)	5.41	2.59, 11.29		
Age Categories (Ordinal)					
18-36 (0), 36-47 (1), >47 (2)	0.4835 (0.23)	1.62	1.14, 2.29		

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

See 2b4.3.

2b4.5. 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*)

1. 100 independent samples of the same size as the original sample were obtained, each of which was a simple random sample with replacement.
2. Logistic regression was applied to each sample using selected risk factors.
3. The 95% confidence intervals based on 100 independent samples for the estimated effects (of the risk factors) were calculated.
4. If the effects at the 2.5th percentile and the 97.5th percentile were both positive (being risk factors) or negative (being protective factors), the effects were deemed to be significant; if the lower and the upper bound of the effects pointed to different directions (one being positive and the other being negative), the effect was deemed to be nonsignificant.
5. Nonsignificant effect was removed from the models, and the stepwise model selection was run to see whether other new effects could enter the models with this effect absent. The above bootstrapping process was repeated to validate the new models.
6. If several effects were found to be nonsignificant through bootstrapping, we removed the least significant effect in step 5.

Table 7. Bootstrap Model Validation Estimates and Corresponding 95% Confidence Intervals, Obtained from 100 Replicates of the Dataset.

Effect	Estimate	95% CI	c-Index	95% CI
ASA Score			0.71	0.67, 0.76
2 vs. 1	0.986	0.534, 2.022		
3/4/5 vs. 1	1.688	1.237,		

		2.867		
Age Categories (Ordinal)				
18-36 (0), 36-47 (1), >47 (2)	0.483	0.060, 1.069		

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

c-index = 0.675

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

Hosmer-Lemeshow p=0.6626

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves: See 2b4.7 for HL statistic.

2b4.9. Results of Risk Stratification Analysis: n/a

2b4.10. 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 model can control for differences in patient case-mix adequately. Further measure maintenance may be required in the future to update the model with more informed and complete datasets.

2b4.11. 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*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE
2b5.1. 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)*

The models calculated the predicted number of surgical site infections. The Standardized Infection Ratio (SIR) and confidence interval were calculated as: reported number of surgical site infections/predicted number of surgical site infections. The SIR is not calculated when the predicted value is less than 0.2. Using the mid-p exact test, the calculated SIR and its confidence interval were compared to an SIR of 1.

2b5.2. 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*)

A meaningful difference in the SIR was defined as an SIR and a confidence interval that was statistically different from 1. Out of 138 total facilities reporting from 2010-2014, SIRs were able to be calculated for 70 of them. Below is a table showing the percentage of SIRs that were significantly different from 1.

Table 8. Distribution of SIRs Calculated for ASCs Reporting from 2010-2014.

SIR	No. of Facilities	Percent
Not Significantly different from 1	60	85.71
Significantly lower than 1	2	2.86
Significantly higher than 1	8	11.43

36 facilities had SIRs not equal to 0.
34 facilities had SIRs equal to 0.
The remaining 68 facilities had a number predicted less than 0.2, so an SIR was not calculated.

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

The SIR enables detection of statistically significant and clinically meaningful differences in SSI that warrant further analysis and possible action. Although exposure volume is low, leading to few statistically significant SIRs in this population, the value of the calculated SIRs can reflect practical measures of performance.

2b6. 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 SDS 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 SDS 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.

2b6.1. 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

2b6.2. 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

2b6.3. 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

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

The crude risks of SSI were compared between the procedures with missing ASA Class and procedures with complete ASA Class using a chi-squared test.

2b7.2. 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*)

There were 8345 missing ASA Classifications out of 46,018 eligible procedures (18.13%). The crude risk in the missing procedures was not significantly different from the crude risk in the included procedures (0.36% compared to 0.25%, p=0.0714).

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

Based on the results above, the missing population does not seem to be significantly different from the included population, minimizing the amount of systemic bias.

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), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

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)*

Some data elements are in defined fields in 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.**

Some data elements are not currently amenable electronic capture, such as physician/nurses notes. NHSN is moving towards complete electronic capture of data as documentation changes occur in ambulatory surgery centers (i.e., as facilities move to full electronic health record capture).

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.**

No feasibility assessment **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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

Use of NHSN surveillance protocol, definitions, and data collection methods for SSI have proven feasible across multiple healthcare settings, including ambulatory surgery centers. Facilities are instructed to follow a standardized data collection procedure (specified by the NHSN protocol and definitions), but specific data collection methods may vary between facilities. Denominator data for breast SSIs are reported as the total number of breast surgical procedures conducted, i.e., an 100% sample. Patient-level data is reported for each procedure and infection; however, the medium of reporting through NHSN is secure and the risk of breaches in patient confidentiality is low. Technical guidance provided by CDC will aid and facilitate accurate data collection and reporting.

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)*.

No fees or licensing requirements. To use this SIR measure, ASCs must be enrolled in NHSN. Detailed instructions on how to enroll can be found here: http://www.cdc.gov/nhsn/ambulatory-surgery/enroll.html.

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.

Planned	Current Use (for current use provide URL)
Payment Program	Public Reporting Colorado Department of Public Health and Environment Patient Safety Program https://www.colorado.gov/pacific/cdphe/hai-surgical-site-infections Public Health/Disease Surveillance National Healthcare Safety Network http://www.cdc.gov/nhsn Quality Improvement with Benchmarking (external benchmarking to multiple organizations) National Healthcare Safety Network http://www.cdc.gov/nhsn Quality Improvement (Internal to the specific organization) National Healthcare Safety Network http://www.cdc.gov/nhsn

4a.1. For each CURRENT use, checked above, provide:

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

Name of Program and Sponsor - National Healthcare Safety Network (NHSN), Centers for Disease Control and Prevention.

Purpose - NHSN is a national system used by CDC and its partners in clinical care and public health for surveillance of healthcare-associated infections, healthcare worker safety, blood safety, antimicrobial use and resistance, and adherence to prevention practices. The system is designed to provide actionable data for healthcare facilities and systems, public health agencies at the state and federal levels, and prevention collaborations. NHSN is the data source for multiple NQF-endorsed measures for which CDC reports measure results on behalf of healthcare facilities to the Centers for Medicare and Medicaid Services (CMS) quality measurement reporting programs.

Geographic area and number and percentage of accountable entities and patients included - NHSN provides national coverage and over 95% of all U.S. hospitals participate in the system. In 2014, there were 435 ASC reporting to NHSN (comprising approximately 8% of all acute care facilities reporting).

Name of program and sponsor - Colorado Department of Public Health and Environment Patient Safety Program

Purpose - Healthcare-associated infections (HAI) are among the top ten leading causes of death in the United States. Colorado recognizes the seriousness of this public health problem and passed the HAI reporting legislation in 2006. House bill 1045 requires hospitals, hospital units, ambulatory surgery centers and dialysis centers to report healthcare-associated infections using the National Healthcare Safety Network (NHSN). This legislation created the Patient Safety Program at the Colorado Department of Public Health and Environment (CDPHE).

Geographic are and number and percentage of accountable entities and patients included - Ambulatory surgery centers (ASCs) began

reporting their measures to NHSN on October 1, 2008. Of the 123 licensed ASCs in the state, 33 provide procedures tracked in NHSN (27%); 33/33 (100%) are currently reporting in NHSN.

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
This is a new measure. Its initial use for public health/disease surveillance, quality improvement with benchmarking (external benchmarking to multiple organizations), and quality improvement (internal to the specific organization) will enable the measure steward, the CDC's National Healthcare Safety Network (NHSN), to identify and address any gaps in the measure specifications that must be closed before CDC can recommend the measure for public reporting or other accountability purposes on the federal level.

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
The CDC's National Healthcare Safety Network (NHSN) will work with ASCs that report SSI data to NHSN to further evaluate the measure's usefulness for SSI prevention and to refine the measure as needed to improve its value for assessing variation in SSI rates intra- and inter-organizationally. NHSN will serve as the data aggregating system. The NHSN Outpatient Procedure Component will provide the technical infrastructure for data collection, analysis, and measure results reporting to participating ASCs, including national benchmarks presented using the SIRs as the summary measures. This additional field experience with measure data, coupled with systematic studies, will serve to define what additional data and methods, if any, are needed to recommend use of this measure for accountability purposes on the federal level.

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

Not applicable

4b.2. **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.**
Not applicable

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
Not applicable

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are

compared to address harmonization and/or selection of the best measure.

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

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 completely harmonized?

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.)
None

Appendix

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

No appendix **Attachment:**

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Surveillance Branch, Division of Healthcare Quality Promotion, Centers for Disease Control and Prevention
Co.2 **Point of Contact:** Daniel, Pollock, dap1@cdc.gov, 404-639-4237-None
Co.3 **Measure Developer if different from Measure Steward:** Surveillance Branch, Division of Healthcare Quality Promotion, Centers for Disease Control and Prevention
Co.4 **Point of Contact:** Daniel, Pollock, dap1@cdc.gov, 404-639-4237-None

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 Development Workgroup:

Donna Slosburg, BSN,LHRM,CASC - Ambulatory Surgery Center Quality Collaboration
Kim Wood, MD - Ambulatory Surgery Center Quality Collaboration
Tamara Hoxworth, Ph.D - Colorado Department of Public Health and Environment
Rosine Angbanzan, MPH - Colorado Department of Public Health and Environment
Katherine Allen-Bridson, RN, BSN, MScPH, CIC - Centers for Disease Control and Prevention
Henrietta Smith, RN, MSN - Centers for Disease Control and Prevention
Janet Brooks, BS, RN, CIC - Centers for Disease Control and Prevention
Rishi Parikh, MPH - Centers for Disease Control and Prevention
Daniel Pollock, MD - Centers for Disease Control and Prevention

ASC Quality Collaboration Technical Expert Committee members:

Naomi Kuznets, PhD - AAAHC
David Shapiro, MD - Ambulatory Surgery Foundation
Gina Throneberry RN, MBA, CASC, CNOR - Ambulatory Surgery Foundation
Kathy Wilson, RN, MHA, LHRM - AMSURG
Linda Brooks-Belli, RN - AMSURG
Jan Davidson, MSN, RN, CNOR, CASC - AORN
Bev Kirchner BSN, CNOR, CASC - AORN
Trey Parsons, RN - ASD Management
Sandra Jones - ASD Management
Melba Willis, RN, BA - Covenant Surgical Partners
Kelly Marcum, BSN, RN - HCA
Sandra Cammon, BS, RN, CPAN, CASC - HCA
Marilyn Parenzan, MBA, RHIA, CPHQ - The Joint Commission
Arwa Muraisi - Kaiser Permanente
Maria Tietjen, RN, BSN, CSPD - OOSS
Lee Anne Blackwell, RN, BSN, CNOR - Practice Partners in Healthcare, Inc
Kathy Bernicky, RN, BSN - Regent Surgical Health
Amiee Mingus, RN, CPAN - Regent Surgical Health
Julie Lewis, BSN, MBA, LHRM - Surgery Partners
Michelle George, RN, MSN, CASC - Surgical Care Affiliates
Ann Shimek, RN, MSN, CASC - USPI
Anita Lambert-Gale, RN - USPI
Amy Glover, BSN, CNOR, CASC - VEI
Patsy Poehler, RN - VEI

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:**
Ad.3 **Month and Year of most recent revision:**
Ad.4 **What is your frequency for review/update of this measure?**
Ad.5 **When is the next scheduled review/update for this measure?**

Ad.6 **Copyright statement:** None
Ad.7 **Disclaimers:** None

Ad.8 **Additional Information/Comments:** None