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 #: 3498e

Measure Title: Hospital Harm- Pressure Injury

Measure Steward: IMPAQ International / CMS

Brief Description of Measure: This electronic clinical quality measure (eCQM) assesses the proportion of inpatient admissions for patients ages 18 years and older who develop a new stage 2, stage 3, stage 4 pressure injury, deep tissue pressure injury, or unstageable pressure injury during hospitalization.

Developer Rationale: This safety eCQM captures the number of patients who experience harm in the form of a pressure injury, during their inpatient hospitalization. Hospital-acquired pressure injuries are serious events and one of the most common patient harms. Pressure injuries commonly cause local infection, osteomyelitis, anemia, and sepsis (Brem, et al., 2010), in addition to causing significant depression, pain, and discomfort to patients (Gunningberg et al., 2011). Pressure injury is considered a serious reportable event by the National Quality Forum (NQF) (Centers for Medicare and Medicaid Services, 2015). CMS also established non-payment for pressure injury (National Quality Forum, 2016), and the rate of pressure injuries is considered an indicator of the quality of nursing care a hospital provides (National Quality Forum, 2005).

It is widely accepted that the risk of developing a pressure injury can be reduced through best practices such as frequent repositioning, proper skin care, and specialized cushions or beds (Berlowitz, et al., 2012). Systematically measuring patients who develop new pressure injuries while in the hospital setting will provide hospitals with a reliable and timely measurement, to more reliably assess harm reduction efforts and modify their improvement efforts in near real-time. This eCQM will fill a gap in measurement and provide incentives for hospitals' quality improvement. Although several pressure injury measures are currently in use, there are no electronic health record (EHR)-based measures intended for use in acute care hospitals. In addition, the intent of this measure is to incentivize greater achievements in reducing harms and enhance hospital performance on patient safety outcomes.

Numerator Statement: The number of hospital inpatient admissions during which a patient developed a new stage 2, stage 3, stage 4 pressure injury, deep tissue pressure injury, or unstageable pressure injury that was not documented as present in the first 24 hours of hospital arrival.

Denominator Statement: All patients 18 years or older at the start of the encounter and discharged inpatient hospital admission during the measurement period. The measure includes inpatient admissions which began in the Emergency Department or in observational status.

Denominator Exclusions: There are no denominator exclusions.

Measure Type: Outcome

Data Source: Electronic Health Records

Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: N/A

Preliminary Analysis: New Measure

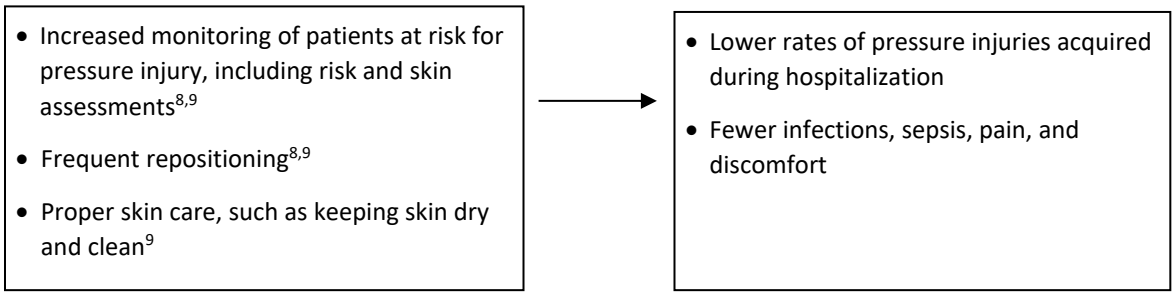
Criteria 1: Importance to Measure and Report

1a. Evidence

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

Evidence Summary

The goal of the Pressure Injury Electronic Clinical Quality Measure (eCQM) is to improve patient safety and prevent patients from acquiring a new pressure injury during their hospitalization. Pressure injuries, also called pressure ulcers, bed sores, or decubitus ulcers, are serious events and one of the most common patient harms. The developer submitted that it is well accepted that pressure injury can be reduced through best practices[8] such as frequent repositioning, proper skin care, and specialized cushions or beds.[3] The desired outcome for this eCQM is a reduction in rates of hospitalized patients who develop a new pressure injury. We define the harm as: a new stage 2, stage 3, stage 4 pressure injury, deep tissue pressure injury, or unstageable pressure injury during hospitalization. The logic model presented by the developer is presented below for this outcome measure.



Question for the Committee:

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

Guidance from the Evidence Algorithm

Outcome measure (Box 1) → Empirical data provided (Box 2) → Pass

RATIONALE:

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.

This eCQM was tested with 3 test sites (24 hospitals) in 3 states (located in Midwest, West, and Northeast). Hospitals varied in size (200+ beds, 15-500 beds, and 450-700 beds), EHR systems (Meditech, Cerner, Epic), teaching status (teaching and non-teaching hospitals), and location (urban, suburban, and rural). A detailed breakdown of the characteristics of the measured facilities and the patient population can be found in the attached Measure Testing Form (Beta Datasets 1, 2, and 3).

The measure performance, including the denominator, numerator, and measure rate by hospital, follows.

Hospital Test Site 1 (Beta Dataset 1 per Testing Form)

- Number of Hospitals: 1
- Data collection period: 1/1/2017 - 12/31/2017
- Denominator: 7,573
- Numerator: 38
- Performance rate: 0.50%
- 95% confidence interval: 0.36%, 0.69%
- Standard Deviation: N/A (only one hospital)

Hospital Test Site 2 (Beta Dataset 2 per Testing Form)

- Number of Hospitals: 21
- Data collection period: 1/1/2017 - 12/31/2017
- Denominator: 100,238
- Numerator: 724
- Performance rate: 0.72%
- 95% confidence interval: 0.67%, 0.78%
- Standard Deviation: 0.47%

Hospital Test 3 (Beta Dataset 3 per Testing Form)

- Number of Hospitals: 2
- Data collection period: 1/1/2017 - 12/31/2017
- Denominator: 56, 330
- Numerator: 414
- Performance rate: 0.73%

- 95% confidence interval: 0.67%, 0.81%
- Standard Deviation: 0.06%

Overall Performance
- Number of Hospitals: 24
- Performance rate: 0.72%
- 95% confidence interval: 0.68%, 0.76%
- Standard deviation: 0.45%
- Range: 0.0% to 1.46%

Disparities

The measure performance was stratified for disparities by age, race, ethnicity, and payer source.

Hospital Test Site 1 (Beta Dataset 1 per Testing Form)
- Number of hospitals: 1
- Data collection period: 1/1/2017 - 12/31/2017
- Denominator (admissions): 7,573

Hospital Test Site 2 (Beta Dataset 2 per Testing Form)
- Number of hospitals: 21
- Data collection period: 1/1/2017 - 12/31/2017
- Denominator (admissions): 100,238

Hospital Test Site 3 (Beta Dataset 3 per Testing Form)
- Number of hospitals: 2
- Data collection period: 1/1/2017 - 12/31/2017
- Denominator (admissions): 56,330

Category//Denominator//Numerator//Measure Rate (95% Confidence Interval)

Across Sites (n=164,141, 24 hospitals)

Age//Denominator//Numerator//Measure Rate (95% Confidence Interval)
18-64//104,332//401//0.38% (0.3%, 0.4%)
65+//59,809//775//1.30% (1.2%, 1.4%)

Gender//Denominator//Numerator//Measure Rate (95% Confidence Interval)
Male//61,636//664//1.08% (1.0%, 1.2%)

Female//102,503//512//0.50% (0.5%, 0.5%)
Unknown//2//0//0.00% (0.0%, 0.7%)

Race//Denominator//Numerator// Measure Rate (95% Confidence Interval)
Black or African American//7,195//51//0.71% (0.5%, 0.9%)
White//133,894//974//0.73% (0.7%, 0.8%)
Other//21,795//142//0.65% (0.5%, 0.8%)
Unknown//1,257//9//0.72% (0.3%, 1.4%)

Ethnicity//Denominator//Numerator//Measure Rate (95% Confidence Interval)
Hispanic or Latino//18,030//89//0.49% (0.4%, 0.6%)
Non-Hispanic//142,251//1,057//0.74% (0.7%, 0.8%)
Unknown//3,860//30//0.78% (0.5%, 1.1%)

(Primary) Payer//Denominator//Numerator// Measure Rate (95% Confidence Interval)
Medicare//64,913//806//1.24% (1.2%, 1.3%)
Medicaid//12,280//96//0.78% (0.6%, 1.0%)
Private Insurance//75,895//236//0.31% (0.3%, 0.4%)
Self-pay or Uninsured//5,999//9//0.15% (0.1%, 0.3%)
Other (such as other government plans)//4,475//27//0.60% (0.4%, 0.9%)
Unknown//579//2//0.35% (0.0%, 1.2%)

It is important to note these results are derived from a small dataset that is not generalizable to the entire population, and the datasets include many characteristics that are 'unknown' in the EHR, which limits the usability of the results.

Questions for the Committee:

- Is there a gap in care that warrants a national performance measure?
- Are there disparities in care that warrant stratification and/or risk adjustment?

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

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

1a. Evidence
Comments:
**Outcomes measure – pass
**appropriate evidence

**The developer provided evidence from AHRQ published data, as well as studies within the last several years that link hospital processes of care to outcome of hospital-acquired pressure injuries.
**This is a new electronic outcome measure. CMS is the sponsor. The desired outcome for this eCQM is a reduction in rates of hospitalized patients who develop a new pressure injury. They define the harm as: a new stage 2, stage 3, stage 4 pressure injury, deep tissue pressure injury, or unstageable pressure injury during hospitalization. The Agency for Healthcare Research and Quality (AHRQ) identified hospital-acquired pressure injuries as a harm to patients that could be prevented, began collecting and reporting incident rates to measure the extent of the problem, and provided toolkits to providers around how to lower their rates.1 It is widely accepted that the risk of developing a pressure injury can be reduced by best practices such as frequent repositioning, proper skin care, and specialized cushions or beds; 2,3 studies have also begun to assess the impact of nutritional interventions. The evidence is not graded; does not include systematic literature reviews.

1b. Performance Gap
Comments:
**Low to moderate – small sample for comparison and may not be generalizable to entire population (3 test sites with 24 hospitals in 3 states); performance rates not statistically different **demonstrated gap
**the measure was tested with 3 beta sites that varied in size, location, EHR systems, and teaching status. Measure performance was stratified for disparities by age, race, ethnicity and payer source. Based on the beta test sites, there appears to be opportunity for improvement, although some members of the expert panel expressed concern about how patient mix could have affected the results while others were concerned about whether documentation deficiencies were affecting the results.
**This eCQM will fill a gap in measurement and provide incentives for hospitals' quality improvement. Although several pressure injury measures are currently in use, there are no electronic health record (EHR)-based measures intended for use in acute care hospitals. In addition, the intent of this measure is to incentivize greater achievements in reducing harms and enhance hospital performance on patient safety outcomes.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

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

Reliability

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

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

and/or that the measure score is precise enough to distinguish differences in performance across providers. For maintenance measures – less emphasis if no new testing data provided.

Validity

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

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

eCQM Technical Advisor(s) review:

Submitted measure is an HQMF compliant eCQM	The submitted eCQMspecifications follow the industry accepted format for eCQM (HL7 Health Quality Measures Format (HQMF)).
Documentation of HQMF,QDM, or CQL limitations	N/A – All components in the measure logic of the submitted eCQM are represented using the HQMF,QDM, or CQL standards;
Value Sets	The submitted eCQM 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. – this includes 100% coverage of measured patient population testing with pass/fail test cases for each population;

Feasibility Testing	Number of data elements included in measure calculation: 11 Number of data elements scoring less than 3 on scorecard: 5 Number of data elements not assessed on scorecard: 2 PressureUlcerStage_Diagnosis • low scoring domains: availability, accuracy, standards, workflow ◦ comments on availability domain: Skin assessments are completed every 12 hours on all patients. We only retained 13% of wound data that had a stage. ◦ comments on accuracy domain: site 1 - Date and Time are captured. When it's documented it is correct. site 2 - Site has had significant variation from a bedside RN documenation and a wound care documentation. If a wound care nurse comes in after an RN has documented something, the max stage may change (increase/decrease), … this may be problematic for the data. ◦ comments on standards domain: Currently built structured data. Could add mapping. ◦ comments on workflow domain: site 1 - Stage 1 pressure injuries may or may not be assessed by the wound nurse. site 2 - Site spent 1 year with this meausre as a goal. Even then, with 10K + nurses there is significant variation. Correct documentaiton of pressure ulcer is complex. Many of our smaller hospitals do not have a wound care specialist (Critical Access). We are working on Telehealth. However, making this part of a national agenda would assist. ObservationServices_EncounterPerformed • Not assessed in scorecard EmergencyDepartmentVisit_EncounterPerformed • not assessed in scorecard

	Diagnosis: Pressure injury date and time • low scoring domains: availability, accuracy, standards, workflow ◦ comments on availability domain: Documented in a "wound" field, making it impossible to distinguish pressure injuries from other wounds. ◦ comments on accuracy domain: site 2 - Only as good as an RN who does a comprehensive assessment. As I understand national standards, it is considered hospital acquired if w/in 24 hours of hospital Admission (MD order to admit). site 3 - POA code often checked off well beyond admission, making identification of POA less reliable unless pressure injury documented in structured field ◦ comments on standards domain: From shift assessment ◦ comments on workflow domain: site 1 - From shift assessment site 3 - Present on admission is documented after 24 hours of admission. Encounter characteristic: Admission date and time • low scoring domains: availability, accuracy, standards, workflow ◦ comments on standards domain: Hard coded by Meditech

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

Evaluators: NQF Scientific Methods Panel Subgroup

Methods Panel Review (Combined)

Evaluation of Reliability and Validity:

Scientific Methods Panel Votes: Measure passes

- Reliability: H-3, M-1, L-1, I-0
- Validity: H-1, M-3, L-0, I-0

This measure was reviewed by the Scientific Methods Panel and discussed on their call. A summary of the measure is provided below:

Reliability

- Reliability was assessed at the measure score level.
- *Note: Data element validity testing was also performed and is discussed in the validity section below. Per NQF guidance, data element validity testing is also acceptable for demonstrating data element reliability.*

- To test reliability of the measure score, the developers used the Adams beta-binomial method (Adams, 2009) to calculate a signal-to-noise ratio.
    - The signal-to-noise ratio yielded a median reliability score of 0.969 (range: 0.206 to 1.000).
- There was some concern among reviewers about the extent to which, in the absence of risk adjustment, the variation between providers represents true variation in quality versus variation due to differences in patient case mix between providers.

Validity

- Testing included both score-level and data element testing.
- Data element
    - Data element validity was assessed by evaluating the accuracy of electronically extracted EHR data elements compared with manually chart abstracted data elements for the same patients, which is considered the "gold standard" for the purpose of these analyses.
    - Positive predictive values (PPV) were calculated for each critical data element:
        - Admission date and time (mm/dd/yyyy hh:mm)—PPV: 100%
        - Diagnosis: Pressure injury date and time—PPV: 94.5%
        - Diagnosis: Pressure injury stage—PPV: 95.6%
        - Patient characteristic: birth date—PPV: 98.2%
- Score Level
    - To demonstrate score-level validity, the developer validated each individual harm identified in a sample of cases in the EHR through chart review by trained abstractors to confirm that the chart, or gold standard, reflects that a harm occurred.
    - Sensitivity, specificity, kappa, and negative predictive value (NPV) were calculated.
    - The developer reports that all but one data element had a match rate of 86% and higher, with most over 91.3%, indicating valid and accurate data elements were extracted from the EHR. The exception was at Beta Dataset 1, pressure injury with date and time, of 72.6%.
    - *Note: Per NQF criteria, the score level testing provided might be more appropriately considered additional data-element validity.*
- There was some concern among reviewers about weak validity results in one of the tested datasets; reviewers suggested that inconsistent use of structured fields in EHRs raises concerns about data quality and documentation practices.

- The developer shared that hospitals were using structured fields, but not documenting in them. One Panel member was concerned that there could be miscategorization based on documentation (or lack thereof) that impacts a hospital's performance. The developer expressed that problematic documentation is part of the quality signal (supported by a TEP) and facilities should be responsible for proper documentation. A panel member responded that one cannot tell if the problem is pressure injury or a documentation issue.

- Some reviewers disagreed with the decision not to risk adjust and/or stratify reported results.

  - Reviewers contended that there are clear differences in patient populations served among hospitals, and suggested risk adjustment or stratification should be considered.

Standing Committee Action Item(s): The Standing Committee can discuss reliability and/or validity or accept the Scientific Methods Panel ratings.

Questions for the Committee regarding reliability:

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

Questions for the Committee regarding validity:

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

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

Scientific Methods Panel Evaluation (Combined): Scientific Acceptability

Scientific Acceptability: Preliminary Analysis Form (Please note multiple answers as answered by multiple Methods Panel members)

Measure Number: 3498e

Measure Title: Hospital Harm- Pressure Injury

Type of measure:

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

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

Data Source:

☐ **Claims** ☒ **Electronic Health Data** ☐ **Electronic Health Records** ☐ **Management Data**

☐ **Assessment Data** ☐ **Paper Medical Records** ☐ **Instrument-Based Data** ☐ **Registry Data**
☐ **Enrollment Data** ☐ **Other**

Level of Analysis:

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

Measure is:

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

RELIABILITY: SPECIFICATIONS

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

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

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

2. **Briefly summarize any concerns about the measure specifications.**
   - Methods Panel Member 1:None.
   - Methods Panel Member 2:Numerator and denominator statements are straight-forward and clear.

RELIABILITY: TESTING

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

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

   ☐ **Yes** ☐ **No** Methods Panel Member 2: (X) NA, score-level reliability was conducted

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

   Methods Panel Member 1: Adams' beta-binomial approach used appropriately.

   Methods Panel Member 2: Score level reliability = beta-binomial method of signal-to-noise (ratio of variances between providers) NOTE: large number of unique patients; relatively small number of hospitals (< 30 across Alpha and Beta testing; measure is reported at facility level.)

   Methods Panel Member 3:SNR based on the beta-binomial model

   Methods Panel Member 4:Calcuated a signal-to-noise ratio.

   Submission document: Testing attachment, section 2a2.2

7. Assess the results of reliability testing

Submission document: Testing attachment, section 2a2.3

Methods Panel Member 1:Based on 164,141 eligible encounters across 24 hospitals in Beta Datasets 1, 2, and 3, the signal-to-noise ratio yielded a median reliability score of 0.969 (range: 0.206 to 1.000), which indicates excellent agreement.

Methods Panel Member 2:Median reliability among 24 hospitals = 0.969, but range = 0.206 – 1.000. Why the very low reliability?

Methods Panel Member 3:Median reliability was excellent (0.97). However, in the absence of risk adjustment, it is unknown to what extent the variation between providers represents true variation in quality versus variation due to differences in patient case mix between providers.

Methods Panel Member 4:Median reliability value was high (0.969).

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

Submission document: Testing attachment, section 2a2.2

☒ **Yes** Methods Panel Member 2: (maybe—extreme low reliability for one or more hospitals is concerning)

☐ **No**

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

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

Submission document: Testing attachment, section 2a2.2

☒ **Yes**

☐ **No**

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

10. OVERALL RATING OF RELIABILITY (taking into account precision of specifications and <u>all</u> testing results):

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

☐ **Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has <u>not</u> been conducted)

☐ **Low** (NOTE: Should rate <u>LOW</u> if you believe specifications are NOT precise, unambiguous, and complete or if testing methods/results are not adequate)

☐**Insufficient** (NOTE: Should rate <u>INSUFFICIENT</u> if you believe you do not have the information you need to make a rating decision)

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

Methods Panel Member 1:Appropriate score level testing was conducted with strong results.

Methods Panel Member 2:Small number of hospitals tested and wide range of reliability scores may be indicative of reliability problems. Is there a data element reliability issue at some of these hospitals even though the data are electronic data?

Methods Panel Member 3:In the absence of risk adjustment, it is unknown to what extent the variation between providers represents true variation in quality versus variation due to differences in patient case mix between providers. It is not enough for MD to simply indicate that "In the case of this hospital-acquired pressure injury eCQM, there is evidence indicating that most newly acquired pressure injuries are avoidable with best practice." The MD needs to provide empiric evidence that patient characteristics, such as age, diabetes, frailty, history of stroke are not associated with the development of pressure ulcers. The high level of score-level reliability may simply reflect the lack of risk adjustment. This is, of course, an empiric question. The MD needs to demonstrate that risk adjustment is not necessary.

Methods Panel Member 4:Score-level testing was conducted; used appropriate method; median reliability value was high.

Methods Panel Member 5: Signal to noise, high

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

Submission document: Testing attachment, section 2b2.

Methods Panel Member 1:None.

Methods Panel Member 2:There are no denominator exclusions for the measure.

Methods Panel Member 3:none

Methods Panel Member 4:Not applicable.

Methods Panel Member 5:N/A

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

Submission document: Testing attachment, section 2b4.

Methods Panel Member 1:None.

Methods Panel Member 2:There was no information/results provided by the developer that showed meaningful differences among providers.

Methods Panel Member 3:The lack of risk adjustment makes it impossible to determine if "measured" differences in performance reflect true differences in quality.

Methods Panel Member 4:No concerns. Variation in hospital performance is noted across the three data sets.

Methods Panel Member 5:No

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

Submission document: Testing attachment, section 2b5.

Methods Panel Member 1:None.

Methods Panel Member 2:Different hospitals apparently have different quality electronic record data. This could (would) cause challenges when comparing results across hospitals.

Methods Panel Member 4:Not applicable.

Methods Panel Member 5:None.

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

- NQF Staff Comment: Missing % for pressure ulcer date / time was 1.2%, and for stage was 1.6%.

Submission document: Testing attachment, section 2b6.

Methods Panel Member 1:None.

Methods Panel Member 2:Developers believe that there may be some missing data. However, given that the measure is not risk adjusted or stratified in reporting, there would be minimal impact of missing data.

Methods Panel Member 4:No concerns. Low frequency of missing data elements.

Methods Panel Member 5:None

16. Risk Adjustment

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

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

☐ Yes ☒ No ☐ Not applicable

16c. **Social risk adjustment:**

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

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

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

16d.**Risk adjustment summary:**

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

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

Methods Panel Member 2: (X) NA—measure is not risk adjusted

16d.3 Is the risk adjustment approach appropriately developed and assessed? ☐ Yes ☐ No
Methods Panel Member 2: (X) NA—measure is not risk adjusted

16d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration)

☐ Yes ☐ No

Methods Panel Member 2: (X) NA—measure is not risk adjusted

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

Methods Panel Member 2: (X) NA—measure is not risk adjusted

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

Methods Panel Member 2:I disagree with the decision to not risk adjust and/or the decision to not stratify the reported results. There are clear differences among patient populations served among hospitals. Patients who are elderly or diabetic or who already have pressure ulcers when admitted have a higher probability of incurring additional pressure ulcers during their stay—inspite of the hospital(s) providing the type of quality care described by the developers. To

compare quality performance across hospitals, some consideration of risk adjustment or stratification should be made.

Methods Panel Member 3:There is no risk adjustment

Methods Panel Member 4:No risk-adjustment, but ok, since literature shows that harm is preventable if best practices are followed.

VALIDITY: TESTING

17. Validity testing level: ☒ Measure score ☒ Data element ☒ Both

18. Method of establishing validity of the measure score:

☐ **Face validity**

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

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

19. Assess the method(s) for establishing validity

Submission document: Testing attachment, section 2b2.2

Methods Panel Member 1:Data element validity testing wsa performed in three "Beta datasets" drawn from 23 hospitals (21 in the second Betaset). From these hospitals, a stratified random sample of total admissions were selected, including 186, 175, and 173 patients respectively. Trained abstrators extracted all of the case information from EMRs at each site and these were compared to the data used to calculate the emeasure.

Methods Panel Member 2:Narrative describing data element validity methodology was confusing (e.g., discussion of "simulating a series of *moe* and target PPV values"). Table 3 presentation of results of methodology provided clearer information.

Narrative describing measure score validity is less confusing, but may slip into discussion of reliability rather than validity. Operational definitions of how sensitivity, specificity, kappa, and negative predicted values were calculated would be useful to display.

The developers argue that since the score is simply the sum harm events, data element validity assures score level validity. They take this a step further by performing measure score level validity testing was performed in a sample of 5 hospitals with a total of 66,127 admissions (the "Alpha dataset") in addition to the three Beta datasets. In this analysis, the fundamental question is whether a patient with a positive result (numerator case) in the EHR data also was a positive result in the abstracted medical record data, as confirmed by a clinical adjudicator, expressed as a positive predictive value (PPV).

Methods Panel Member 3:The agreement between EMR and chart was tested by re-abstraction.

Methods Panel Member 4:Measure score level testing: compared if events captured by the e-measure matched events captured by manual abstraction.

Data element testing: compared electronic abstracted values to manual abstracted values.

Both approaches seem reasonable.

20. Assess the results(s) for establishing validity

Submission document: Testing attachment, section 2b2.3

Methods Panel Member 1:At the data element level, all but one data element had a match rate of 86% and higher, with most over 91.3%, indicating valid and accurate data elements were extracted from the EHR. Minor discrepancies were explained.

At the score level, PPV was high in the Alpha Dataset and Beta Dataset 2, at 97.8% and 97%, respectively, meaning that in almost all cases, the admissions met the criteria for a harm in the chart abstracted and EHR data. Beta Dataset 1 and Beta Dataset 3 had lower PPV at 68.4% and 44.9% respectively, which the developers explain as documentation errors. They conclude that the measure is highly valid when hospitals consistently use structured fields to document pressure injuries.

Methods Panel Member 2:

Table 3 seems to show reasonable results for data element validity.

Table 4 seem to show reasonable results. However, only Beta dataset 2 has very strong results, with Beta dataset 3 results being quite a bit lower than either of the previous sets of results. Why? Explanations that were offered seem to be related to small sample size issues where a few errors lead to poor results. The other explanation (poor documentation practices) call into question the overall quality of electronic health records and may be problematic when the measure is applied nationally.

Methods Panel Member 3:Assessed the validity of outcome data element using sensitivity, specificity, NPV, and kappa statistic. These measures of agreement were good for 2 of the 3 test data sets.

Methods Panel Member 4:Measure score validity: Two of the data sets had lower PPVs (68.4% and 44.9%). Concerns with documentation practices and not putting data into structure fields.

Data element validity: One of the data elements had a PPV of 72.6% in one of the beta data sets. But it had higher PPVs in the other beta data sets.

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

Submission document: Testing attachment, section 2b1.

☒ **Yes**

☐ **No**

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

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

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

☒ **Yes**

☐ **No**

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

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

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

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

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

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

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

Methods Panel Member 1:My primary concern is that the score level analysis showed that the measure is highly valid only when hospitals consistently use structured fields to document pressure injuries.

Methods Panel Member 2:Score validity could be problematic as the underlying quality of electronic health records was called into question by the developer's explanation for why two of the Beta tests showed poor results.

Methods Panel Member 3:The lack of risk adjustment is a critical limitation of this measure. A priori, it would be difficult to assume that patient frailty and comorbidities do not play an important role in the development of pressure ulcers. There is clearly a spectrum of risk between young patients who are able to ambulate and stroke patients who spend much of their day in bed or in a chair.

Methods Panel Member 5:Data element: compare EHR with charts, PPV

Score: sensitivity, specificity, kappa and NPV

Methods Panel Member 4:Concerns with the inconsistent use of structured fields by hospitals and how that inconsistency might influence an individual hospital's score on the measure (i.e., the hospital might look like they had no events, but actually did).

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

2a1. Reliability – Specifications
Comments:
**High – Scientific Methods Panel passes the measure with reliability votes: H-3, M-1, L-1, I-0; reliability at the measure score level: 0.969
**None
**The PSC should discuss the lack of case-mix adjustment as a potential factor affecting reliability
**Beta Dataset 1 did not have a field for pressure injuries in their EHR, only one for "wounds"; this meant we only used pressure injury data elements that had a stage attached, to ensure all data used were pressure injuries and not extraneous wounds. This health system is remedying their EHR currently. Beta Datasets 2 and 3 had extremely low rates of missing data required for the measure calculation. We looked at the pressure injury-level instead of admission-level to get a clearer picture of the data reliability. The findings indicate that for health systems that are able to identify pressure injuries in discrete fields, all data elements required to calculate the measure are readily available and infrequently missing.

2a2. Reliability – Testing
Comments:

**No concerns, note no risk adjustment to explain any difference between quality vs. patient mix of two providers
**No concerns
**No
**More testing will be done in the future.

2b1. Validity –Testing
Comments:
**Moderate – Scientific Methods Panel passes the measure with validity votes: H-1, M-3, L-0, I-0; validity testing at score level and data element level: positive predictive value (PPV) of 94.5%-100% for data element; PPV of >86% at score level except one outliner at 72.6%
**No concerns
**No
**Detail was provided

2b4-7. Threats to Validity
2b4. Meaningful Differences
Comments:
**No concerns; the submitted eCQMspecifications follow the industry accepted format for eCQM (HL7 Health Quality Measures Format (HQMF))
**None
**Other than the concern about whether case-mix adjustment of the measure should be considered, I have no concerns
**NA not in use yet

2b2-3. Other Threats to Validity
2b2. Exclusions
2b3. Risk Adjustment
Comments:
**Data not risk adjusted, consider risk adjusting
**agree there should be some stratification - bed size, teaching vs non teaching etc
**See previous comments regarding possible need for case-mix adjustment
**No risk adjustment or risk stratification

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 are generated or collected by and used by healthcare personnel during the provision of care and originate in Electronic Health Records
- During the testing of the eCQM, it was demonstrated that extraction of this measure reliability from the EHR for several data elements is not feasible.

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 eCQM, does the eCQM Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites?

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

RATIONALE:

- Data are readily available from electronic health records and are captured in the process of delivering care.
- As noted above, feasibility assessment across 24 hospitals with three different EHR vendors showed that most data elements used to calculate the measure were reliably available in a structured format within the EHR, captured as part of the course of care, and coded using nationally accepted terminology. However, during testing by NQF, this was not validated and several data elements demonstrated problems with feasibility.

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

3. Feasibility
Comments:
**Low to moderate – data collected via electronic health records, but extraction for several data elements not feasible
**concerns about feasibility from NQF staff an area for discussion
**No concerns
**This eCQM was tested with 3 test sites (24 hospitals) in 3 states (located in Midwest, West, and Northeast). Hospitals varied in size (200+ beds, 15-500 beds, and 450-700 beds), EHR systems (Meditech, Cerner, Epic), teaching status (teaching and non-teaching hospitals), and location (urban, suburban, and rural).

Criterion 4: Usability and Use

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

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

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

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

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

OR

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

Accountability program details

- This eCQM is under initial endorsement review and is not currently used in any accountability program. In December 2018, this eCQM was presented to the Measure Applications Partnership (MAP), who noted conditional support for rulemaking. Thus, the MAP is recommending implementation in an accountability program pending feedback received during NQF endorsement and rulemaking.

4a.2. Feedback on the measure by those being measured or others. N/A

Feedback on the measure by those being measured or others

- According to the developer, while this measure does not have usability information from measured entities, as it is being developed de novo and has not been implemented yet, the development eam team sought input from multiple stakeholder groups throughout the measure development process. During development, a technical expert panel composed of a variety of stakeholders was engaged at various stages of development to obtain balanced, expert input. The developer also solicited and received feedback on the measure through an MMS Blueprint 44-day Public Input Period during development.
- Developer states that input received from TEP members was instrumental to the development and specification of this measure. Feedback received during public comment was also explored during the measure testing process.

Additional Feedback: N/A

***Questions for the Committee**:*

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

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

RATIONALE:

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

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

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

Improvement results N/A

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

Unexpected findings (positive or negative) during implementation N/A

Potential harms N/A

Additional Feedback: N/A

***Questions for the Committee**:*

- 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**

RATIONALE:

- Other pressure ulcer measures exist that are used in public programs.

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

4a1. Use - Accountability and Transparency

Comments:
**New measure not used in any accountability program
**concerned about use in public reporting when this is not a feasible measure to capture
**Not currently used in accountability program, but CMS plans to consider this
**This is a new eCQM and there is no time trend information available regarding facility performance improvement. This eCQM is not currently used in any quality improvement program, but a primary goal of the eCQM is to provide hospitals with performance information necessary to implement focused quality improvement efforts. CMS also established non-payment for pressure injury (National Quality Forum, 2016), and the rate of pressure injuries is considered an indicator of the quality of nursing care a hospital provides (National Quality Forum, 2005). It is widely accepted that the risk of developing a pressure injury can be reduced through best practices such as frequent repositioning, proper skin care, and specialized cushions or beds (Berlowitz, et al., 2012). Systematically measuring patients who develop new pressure injuries while in the hospital setting will provide hospitals with a reliable and timely measurement, to more reliably assess harm reduction efforts and modify their improvement efforts in near real-time. This eCQM will fill a gap in measurement and provide incentives for hospitals' quality improvement. Although several pressure injury measures are currently in use, there are no electronic health record (EHR)-based measures intended for use in acute care hospitals. In addition, the intent of this measure is to incentivize greater achievements in reducing harms and enhance hospital performance on patient safety outcomes.

4b1. Usability – Improvement

Comments:
**input from technical expert panel, reduce patient harm
**i think this is useful however there are other measures out there that capture HAPU -- not to mention concerned about feasibility
**I believe the usability of the measure is high and could further the goal of prevention of hospital-acquired pressure injuries

**While there are several measures that target the reduction of hospital-acquired pressure injuries in use in various patient populations, there are no eCQMs intended for use to compare quality across acute care hospitals. The measures NQF# 0679 and #0678 target a different patient population and use chart review data from the following sources: Minimum Data Set (MDS); Long Term Care Hospitals Continuity Assessment Record and Evaluation (LTCH-CARE) Data set; and the Inpatient Rehabilitation Facility Patient Assessment Instrument (IRF-PAI) Data set. Additionally, NQF# 0678 measure includes worsening pressure injuries and NQF# 0679's population consists of only high-risk patients defined as those who are impaired in bed mobility, comatose, or suffering malnutrition. The new Hospital Harm -Pressure Injury eCQM identifies pressure injuries using direct extraction of structured data from the EHR and will provide hospitals with reliable and timely measurement of their pressure injury rates. As these measures do not apply to the same measured entities, it should not impact data collection burden. Final measure specifications for implementation will be made publicly available on CMS' appropriate quality website, once finalized through the NQF endorsement and CMS rulemaking processes.

Criterion 5: Related and Competing Measures

Related or competing measures

- Competing: Hospital-acquired pressure injuries are currently measured and publicly reported in the Hospital-Acquired Condition Reduction Program (HACRP) as a component of the Patient Safety Indicator (PSI) 90 measure, which relies on ICD codes as a data source.
- Related: Additionally, the following NQF endorsed measures are related but measure different patient populations: Percent of High Risk Residents with Pressure Ulcers (Long Stay) (NQF #0679) and Percent of Residents or Patients with Pressure Ulcers that are New or Worsened (Short Stay) (NQF #0678).

Harmonization

- According to the developer:
    - While there are several measures that target the reduction of hospital-acquired pressure injuries in use in various patient populations, there are no eCQMs intended for use to compare quality across acute care hospitals. The measures NQF# 0679 and #0678 target a different patient population and use chart review data from the following sources: Minimum Data Set (MDS); Long Term Care Hospitals Continuity Assessment Record and Evaluation (LTCH-CARE) Data set; and the Inpatient Rehabilitation Facility Patient Assessment Instrument (IRF-PAI) Data set. Additionally, NQF# 0678 measure includes worsening pressure injuries and NQF# 0679's population consists of only high-risk patients defined as those who are impaired in bed mobility, comatose, or suffering malnutrition. The new Hospital Harm -Pressure Injury eCQM identifies pressure injuries using direct extraction of structured data from the EHR and will provide hospitals with reliable and timely measurement of their pressure injury rates. As these measures do not apply to the same measured entities, it should not impact data collection burden.
    - Hospital-acquired pressure injuries are currently measured and publicly reported in the Hospital-Acquired Condition Reduction Program (HACRP) as a component of the Patient Safety Indicator (PSI) 90 measure (PSI-03). PSI-03 does not include stage 2 pressure

injuries in the outcome, has additional exclusions to the cohort, and uses ICD codes via claims as a data source. Hospital Harm - Pressure Injury Measure is an eCQM (EHR data-only), which stakeholders and TEP have noted as a more desirable data source with more face validity for measuring pressure injuries.

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

5. Related and Competing

Comments:

**Competing measures on hospital acquired pressure injuries; related measures #0679 Percent of High Risk Residents with Pressure Ulcers (Long Stay) and #0678 (Percent of Residents or Patients with Pressure Ulcers that are New or Worsened (Short Stay)); this measure is an e-measure and includes stage 2 pressure injuries; suggest harmonizing measures

**i am concerned about other measures as well as NDNQI measures for pressure injuries

**There are related and competing measures, but the developer discussed how they differ from this measure. I am satisfied with their explanation

**Competing: Hospital-acquired pressure injuries are currently measured and publicly reported in the Hospital-Acquired Condition Reduction Program (HACRP) as a component of the Patient Safety Indicator (PSI) 90 measure, which relies on ICD codes as a data source. Related: Additionally, the following NQF endorsed measures are related but measure different patient populations: Percent of High Risk Residents with Pressure Ulcers (Long Stay) (NQF #0679) and Percent of Residents or Patients with Pressure Ulcers that are New or Worsened (Short Stay) (NQF #0678).

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 6/5/2019

- **No NQF Members have submitted support/non-support choices as of this date.**

Public Comment

**Importance: According to AHRQ Partnership For Patients' Program, pressure injuries are the second most common adverse event behind drug events. Thus, having a usable metric for this patient safety event is imperative. It needs to be stressed this is an important "replacement metric" closing an important patient safety measurement "gap," since the impact of the current PSI-90 pressure injury metric (PSI 03) has been mitigated due to concerns regarding its use of administrative data and its validity.

Pressure Injury should be viewed as 100% preventable and aggressive preventative strategies should be implemented in all at-risk patients, not just those showing signs of impending ulcers. These include, mattress cushions, turning the patient every 2 hours and preemptively padding areas which are prone to form ulcers. Thus, whether or not a Stage I injury is present, prompt preventative strategies on all at-risk patients should prevent progression in the vast majority of patients.

Advantages of the Replacement Metric: One of the major advantages of the proposed metric is that it utilizes EMR and not Administrative Billing Data. The latter has long been held by the industry as having a

low validity. In addition, the definition of the metric has been changed. It now measures injury with any skin breakdown (Stage II, III, and IV pressure injuries), avoiding a subjective judgement on the depth of the ulcer. Thus, when drainage is observed or when there is lack of skin integrity an event will be captured. (Note: Stage I injury is a discoloration of skin without skin breakdown).
The current PSI 03 metric only reports Stage III and IV pressure injuries, which when entering data into the EMR requires a subjective judgement on depth in the differentiation of Stage II and Stage III. Such a judgement would be expected to require additional training and the metric would be expected to have decreased validity and reliability. In addition, it does not measure all pressure ulcers, since Stage 2 ulcers are not captured.
Burden: There should be little burden on the facility, since the EMR systems can be used to captures the events. Thus, the burden should be similar to that of the original PSI 03 metric.
Disparities: Disparities is an important issue. In pressure injuries, healthcare resources and socioeconomic factors are of paramount importance and should not be mathematically negated but instead corrected. Stage II, III and IV pressure ulcers which are present on or develop within 24 hours of admission are captured. The 24 hour grace period will allow for identification of latent pressure injury. This should correct for preadmission ulcer formation caused by access and socioeconomic disparities. In a study of nursing home residents, Park Lee, et al, in a NCHS Data Brief reviewed over 159,000 nursing home residents and found that "Pressure ulcer prevalence varied by age, sex, and length of time since admission to the nursing home, but not by race." https://www.cdc.gov/nchs/data/databriefs/db14.pdf

**The Public Policy Committee and the Board of Directors of the National Pressure Ulcer Advisory Panel (NPUAP), are reaching out to you in response to the open comment period for Measures #3498e titled "Hospital Harm Pressure Injury".

The NPUAPis an independent, not-for-profit professional organization dedicated to the prevention and management of pressure injuries. Formed in 1987, the NPUAP Board of Directors is composed of leading experts from diverse health care disciplines—all of whom share a commitment to the prevention and management of pressure injuries. The NPUAP serves as a resource to health care professionals, government, the public, and health care agencies. The NPUAP welcomes and encourages the participation of those interested in pressure injury issues through the utilization of NPUAP educational materials, participation at national conferences, and support of efforts in public policy, education and research.

The NPUAP suggests that further clarification, research and/or edits for this measure would be beneficial pertaining to the following points:

• Proposed 24-hour time frame from admission to declare a hospital acquired pressure injury is not consistent with current science. ◦As the science surrounding the evolution of a Deep Tissue Pressure Injury (DTPI) continues to advance, it has been postulated that the appearance of a DTPI can take up to 48 hours or longer to manifest and become visible to the clinician. Therefore, a 24 hour timeframe to declare a pressure injury (specifically a deep tissue pressure injury) as hospital acquired may erroneously penalize institutions for pressure injuries that may have developed prior to admission, but are not visible to clinicians within 24 hours of admission.

◦Moreover, current and emerging technologies such as the use of infrared thermographic devices, ultrasound and subepidermal moisture devices support that changes in tissues may be developing below

the skin surface, and before visible signs are present to the clinician. Thus, there are some pressure injuries that may actually be present on admission, however not visible within the first 24 hours.

◦Similarly, in darker pigmented skin, it may be difficult to visualize a potential deep tissue injury or Stage 1 pressure injury in its early stages, which can also contribute to the erroneous labelling of a hospital acquired pressure injury in these individuals, as skin changes may not be readily detected within the first 24 hours of the hospital admission.

◦Based on these clinical concerns, the NPUAP strongly believes that reconsideration for this 24 hour timeframe should be undertaken. A suggestion might be to have an algorithm that states Stage 2, 3, 4 & unstageable pressure injuries should be documented within 24 hours of admission. In the case of a DTPI, a 48 hour time frame or longer could be proposed in which the clinician would document the presence of a DTPI.

• The proposed e-measure lacks clear guidance as to where in the EMR the pressure injury documentation will be extracted. ◦It is unclear from the proposed measure where the information on pressure injury development to support the label of a hospital acquired pressure injury would be obtained within the EMR. In many EMR systems, there are multiple places to document a similar finding, leading to confusion and inconsistencies. This concern was supported by comments from the Meditech users in your beta site testing, who stated "documented in the wound field, making it impossible to distinguish a pressure injury from another type of wound."

◦ Furthermore, it is unclear if this information will be extracted from a nursing flowsheet, admission assessment or from the provider/midlevel practitioner in free texted notes. Caution has been recommended when interpreting data from an operational EMR, as data inaccuracy, incompleteness or missing data are all consequences of the use of an EMR. (Hersh et al., 2013). Varied descriptions of data elements across multiple EMR vendors, variability in documentation style and multiple locations within the EMR in which to document clinical events such as pressure injuries all contribute to ambiguity in data interpretation.

◦The proposed measure lacks clear direction as to the location in the EMR the stage of pressure injury will be pulled. Accurate staging of pressure injuries has been a concern for decades and this concern crosses all disciplines. Studies evaluating clinician knowledge of pressure injury staging using a standardized tool have found that nurses consistently score in the "C" to "C+" range with similar results for physicians. While some facilities allow RNs to stage pressure injuries, others do not. Lack of the availability of a wound care clinician to corroborate or assign a pressure injury stage can lead to erroneous staging, thus inaccurate documentation. Institutions that lack wound care clinicians will be placed at a clear disadvantage as a result of this proposed measure. These concerns are corroborated with your beta test sites as it was noted that there was difficulty determining pressure injury stage from the documentation and concerns were raised regarding the accuracy of the pressure injury staging, especially in hospitals that did not have the availability of a wound care clinician to determine the stage the pressure injury.

• The NPUAP has concerns related to the validity and reliability of the proposed measure based on the scorecard results provided and previous experiences in developing pressure injury e-measures (Warren & Dunton, 2014). ◦Overall, according to the summary scorecard, data accuracy for pressure injury date and time was identified as 0% and pressure injury stage was identified at 33%. The reliability and validity of the information extracted for this proposed measure is therefore a concern. It is clear that there remains much work to be done across the United States with respect to the accuracy of pressure injury staging and documentation before an e-measure such as the one proposed can be initiated.

◦At the NPUAP, one of our primary goals is to provide pressure injury education to all disciplines, across all types of health care settings and perhaps this issue warrants more attention on a national level for which the NPUAP could be a lead partner.

The NPUAP would be happy to continue our ongoing collaboration with the NQF and CMS to support the educational needs associated with the full understanding of these terms and measures necessary for accurate clinical classification/staging. Thank you for the opportunity to comment.

**(2nd comment by individual at same agency) The Public Policy Committee and the Board of Directors of the National Pressure Ulcer Advisory Panel (NPUAP), are reaching out to you in response to the open comment period for Measures #3498e titled "Hospital Harm Pressure Injury".

The NPUAPis an independent, not-for-profit professional organization dedicated to the prevention and management of pressure injuries. Formed in 1987, the NPUAP Board of Directors is composed of leading experts from diverse health care disciplines—all of whom share a commitment to the prevention and management of pressure injuries. The NPUAP serves as a resource to health care professionals, government, the public, and health care agencies. The NPUAP welcomes and encourages the participation of those interested in pressure injury issues through the utilization of NPUAP educational materials, participation at national conferences, and support of efforts in public policy, education and research.

The NPUAP suggests that further clarification, research and/or edits for this measure would be beneficial pertaining to the following points:

• Proposed 24-hour time frame from admission to declare a hospital acquired pressure injury is not consistent with current science. ◦As the science surrounding the evolution of a Deep Tissue Pressure Injury (DTPI) continues to advance, it has been postulated that the appearance of a DTPI can take up to 48 hours or longer to manifest and become visible to the clinician. Therefore, a 24 hour timeframe to declare a pressure injury (specifically a deep tissue pressure injury) as hospital acquired may erroneously penalize institutions for pressure injuries that may have developed prior to admission, but are not visible to clinicians within 24 hours of admission.

◦Moreover, current and emerging technologies such as the use of infrared thermographic devices, ultrasound and subepidermal moisture devices support that changes in tissues may be developing below the skin surface, and before visible signs are present to the clinician. Thus, there are some pressure injuries that may actually be present on admission, however not visible within the first 24 hours.

◦Similarly, in darker pigmented skin, it may be difficult to visualize a potential deep tissue injury or Stage 1 pressure injury in its early stages, which can also contribute to the erroneous labelling of a hospital acquired pressure injury in these individuals, as skin changes may not be readily detected within the first 24 hours of the hospital admission.

◦Based on these clinical concerns, the NPUAP strongly believes that reconsideration for this 24 hour timeframe should be undertaken. A suggestion might be to have an algorithm that states Stage 2, 3, 4 & unstageable pressure injuries should be documented within 24 hours of admission. In the case of a DTPI, a 48 hour time frame or longer could be proposed in which the clinician would document the presence of a DTPI.

• The proposed e-measure lacks clear guidance as to where in the EMR the pressure injury documentation will be extracted. ◦It is unclear from the proposed measure where the information on pressure injury development to support the label of a hospital acquired pressure injury would be obtained within the EMR. In many EMR systems, there are multiple places to document a similar finding, leading to confusion and inconsistencies. This concern was supported by comments from the Meditech users in your beta site testing, who stated "documented in the wound field, making it impossible to distinguish a pressure injury from another type of wound."

◦ Furthermore, it is unclear if this information will be extracted from a nursing flowsheet, admission assessment or from the provider/midlevel practitioner in free texted notes. Caution has been recommended when interpreting data from an operational EMR, as data inaccuracy, incompleteness or missing data are all consequences of the use of an EMR. (Hersh et al., 2013). Varied descriptions of data elements across multiple EMR vendors, variability in documentation style and multiple locations within the EMR in which to document clinical events such as pressure injuries all contribute to ambiguity in data interpretation.

◦The proposed measure lacks clear direction as to the location in the EMR the stage of pressure injury will be pulled. Accurate staging of pressure injuries has been a concern for decades and this concern crosses all disciplines. Studies evaluating clinician knowledge of pressure injury staging using a standardized tool have found that nurses consistently score in the "C" to "C+" range with similar results for physicians. While some facilities allow RNs to stage pressure injuries, others do not. Lack of the availability of a wound care clinician to corroborate or assign a pressure injury stage can lead to erroneous staging, thus inaccurate documentation. Institutions that lack wound care clinicians will be placed at a clear disadvantage as a result of this proposed measure. These concerns are corroborated with your beta test sites as it was noted that there was difficulty determining pressure injury stage from the documentation and concerns were raised regarding the accuracy of the pressure injury staging, especially in hospitals that did not have the availability of a wound care clinician to determine the stage the pressure injury.

• The NPUAP has concerns related to the validity and reliability of the proposed measure based on the scorecard results provided and previous experiences in developing pressure injury e-measures (Warren & Dunton, 2014). ◦Overall, according to the summary scorecard, data accuracy for pressure injury date and time was identified as 0% and pressure injury stage was identified at 33%. The reliability and validity of the information extracted for this proposed measure is therefore a concern. It is clear that there remains much work to be done across the United States with respect to the accuracy of pressure injury staging and documentation before an e-measure such as the one proposed can be initiated.

◦At the NPUAP, one of our primary goals is to provide pressure injury education to all disciplines, across all types of health care settings and perhaps this issue warrants more attention on a national level for which the NPUAP could be a lead partner.

The NPUAP would be happy to continue our ongoing collaboration with the NQF and CMS to support the educational needs associated with the full understanding of these terms and measures necessary for accurate clinical classification/staging. Thank you for the opportunity to comment.

Brief Measure Information

NQF #: 3498e

Corresponding Measures:

De.2. Measure Title: Hospital Harm - Pressure Injury

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

De.3. Brief Description of Measure: This electronic clinical quality measure (eCQM) assesses the proportion of inpatient admissions for patients ages 18 years and older who develop a new stage 2, stage 3, stage 4 pressure injury, deep tissue pressure injury, or unstageable pressure injury during hospitalization.

1b.1. Developer Rationale: This safety eCQM captures the number of patients who experience harm in the form of a pressure injury, during their inpatient hospitalization. Hospital-acquired pressure injuries are serious events and one of the most common patient harms. Pressure injuries commonly cause local infection, osteomyelitis, anemia, and sepsis (Brem, et al., 2010), in addition to causing significant depression, pain, and discomfort to patients (Gunningberg et al., 2011). Pressure injury is considered a serious reportable event by the National Quality Forum (NQF) (Centers for Medicare and Medicaid Services, 2015). CMS also established non-payment for pressure injury (National Quality Forum, 2016), and the rate of pressure injuries is considered an indicator of the quality of nursing care a hospital provides (National Quality Forum, 2005).

It is widely accepted that the risk of developing a pressure injury can be reduced through best practices such as frequent repositioning, proper skin care, and specialized cushions or beds (Berlowitz, et al., 2012). Systematically measuring patients who develop new pressure injuries while in the hospital setting will provide hospitals with a reliable and timely measurement, to more reliably assess harm reduction efforts and modify their improvement efforts in near real-time. This eCQM will fill a gap in measurement and provide incentives for hospitals' quality improvement. Although several pressure injury measures are currently in use, there are no electronic health record (EHR)-based measures intended for use in acute care hospitals. In addition, the intent of this measure is to incentivize greater achievements in reducing harms and enhance hospital performance on patient safety outcomes.

References:

Brem H, M. J., Nierman D, et al. (2010). High Cost of Stage IV Pressure Ulcers. doi:10.1016/j.amjsurg.2009.12.021. American Journal of Surgery, 200(4), 473-477.

Gunningberg, L., Donaldson, N., Aydin, C., Idvall, E. (2011). Exploring variation in pressure ulcer prevalence in Sweden and the USA: Benchmarking in action. 18. 10.1111/j.1365-2753.2011.01702.x. Journal of evaluation in clinical practice., 904-910.

Centers for Medicare & Medicaid Services. (2015). Hospital-Acquired Conditions. Retrieved January 13, 2017, from https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HospitalAcqCond/Hospital-Acquired_Conditions.html

National Quality Forum. (2016). List of SREs. Retrieved January 13, 2017, from http://www.qualityforum.org/Topics/SREs/List_of_SREs.aspx#sre4

National Quality Forum. (2005). National Voluntary Consensus Standards for Nursing-Sensitive Care: An Initial Performance Measure Set. Retrieved January 13, 2017, from http://www.qualityforum.org/Publications/2004/10/National_Voluntary_Consensus_Standards_for_Nursing-Sensitive_Care__An_Initial_Performance_Measure_Set.aspx.

Berlowitz, D. VanDeusen Lukas, C.; Parker, V.; Niederhauser, A.;, & Silver, J. L., C.; Ayello, E.; Zulkowski, K. (2012). Preventing Pressure Ulcers in Hospitals- A Toolkit for Improving Quality of Care.

S.4. Numerator Statement: The number of hospital inpatient admissions during which a patient developed a new stage 2, stage 3, stage 4 pressure injury, deep tissue pressure injury, or unstageable pressure injury that was not documented as present in the first 24 hours of hospital arrival.

S.6. Denominator Statement: All patients 18 years or older at the start of the encounter and discharged inpatient hospital admission during the measurement period. The measure includes inpatient admissions which began in the Emergency Department or in observational status.

S.8. Denominator Exclusions: There are no denominator exclusions.

De.1. Measure Type: Outcome

S.17. Data Source: Electronic Health Records

S.20. 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? N/A

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

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

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

FORCMS_Hospital_Harm_PI_NQF_Evidence_Form.docx

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

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

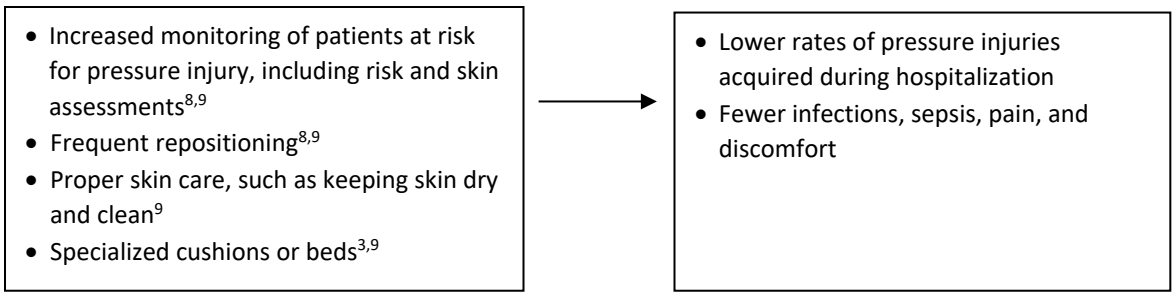
No

[NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** N/A
Measure Title: Hospital Harm- Pressure Injury
IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: N/A
Date of Submission: 4/2/2019
1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)
Outcome
☒ Outcome: Hospital Harm – Pressure Injury
☐Patient-reported outcome (PRO): Click here to name the PRO
PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors. (*A PRO-based performance measure is not a survey instrument. Data may be collected using a survey instrument to construct a PRO measure.)*
☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome
☐ Process: Click here to name what is being measured
☐ Appropriate use measure: Click here to name what is being measured
☐ Structure: Click here to name the structure
☐ Composite: Click here to name what is being measured

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

The goal of the Pressure Injury Electronic Clinical Quality Measure (eCQM) is to improve patient safety and prevent patients from acquiring a new pressure injury during their hospitalization. Pressure injuries, also called pressure ulcers, bed sores, or decubitus ulcers, are serious events and one of the most common patient harms. The injury can present as intact skin or an open ulcer, may be painful, and occurs from unrelieved pressure on the skin or in combination with shear force. Pressure injuries commonly lead to further patient harm, including local infection, osteomyelitis, anemia, and sepsis, in addition to causing significant depression, pain, and discomfort to patients.[1,2,3] The presence or development of a pressure injury can increase the length of a patient's hospital stay by an average of four days, which increases spending ranging from $20,900 to $151,700 per pressure injury.[4] Pressure injury is considered a serious reportable event by the National Quality Forum (NQF),[5] the CMS established non-payment for pressure injury,[6] and it is considered an indicator of the quality of nursing care a hospital provides.[7] It is well accepted that pressure injury can be reduced through best practices[8] such as frequent repositioning, proper skin care, and specialized cushions or beds.[3] The desired outcome for this eCQM is a reduction in rates of hospitalized patients who develop a new pressure injury. We define the harm as: a new stage 2, stage 3, stage 4 pressure injury, deep tissue pressure injury, or unstageable pressure injury during hospitalization.


• Increased monitoring of patients at risk for pressure injury, including risk and skin assessments[8,9]
• Frequent repositioning[8,9]
• Proper skin care, such as keeping skin dry and clean[9]
• Specialized cushions or beds[3,9]
• Lower rates of pressure injuries acquired during hospitalization
• Fewer infections, sepsis, pain, and discomfort

References:

1. Brem H MJ, Nierman D, et al. High Cost of Stage IV Pressure Ulcers. doi:10.1016/j.amjsurg.2009.12.021. American Journal of Surgery. 2010;200(4):473-477.
2. National Pressure Ulcer Advisory Panel. NPAUAP Pressure Injury Stages 2016; http://www.npuap.org/resources/educational-and-clinical-resources/npuap-pressure-injury-stages/.
3. Gunningberg L, Donaldson, N., Aydin, C., Idvall, E. Exploring variation in pressure ulcer prevalence in Sweden and the USA: Benchmarking in action. 18. 10.1111/j.1365-2753.2011.01702.x. Journal of evaluation in clinical practice. 2011: 904-910.
4. Bauer K, Rock K, Nazzal M, Jones O, Qu W. Pressure Ulcers in the United States' Inpatient Population From 2008 to 2012: Results of a Retrospective Nationwide Study. Ostomy Wound Manage. 2016;62(11):30-38.
5. National Quality Forum. List of SREs. 2016; http://www.qualityforum.org/Topics/SREs/List_of_SREs.aspx#sre4. Accessed January 13, 2017.
6. Centers for Medicare & Medicaid Services. Hospital-Acquired Conditions. 2015; https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HospitalAcqCond/Hospital-Acquired_Conditions.html. Accessed January 13, 2017.
7. National Quality Forum. National Voluntary Consensus Standards for Nursing-Sensitive Care: An Initial Performance Measure Set. Washington, D.C.: National Quality Forum; 2004.
8. Agency for Healthcare Research and Quality. Preventing Pressure Ulcers in Hospitals - A Toolkit for Improving Quality of Care. 2012; Rockville, MD. http://www.ahrq.gov/professionals/systems/hospital/pressureulcertoolkit/index.html
9. National Pressure Ulcer Advisory Panel, European Pressure Ulcer Advisory Panel, Pan Pacific Pressure Injury Alliance. Prevention and Treatment of Pressure Ulcers: Quick Reference Guide. 2014.

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

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

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

The Agency for Healthcare Research and Quality (AHRQ) identified hospital-acquired pressure injuries as a harm to patients that could be prevented, began collecting and reporting incident rates to measure the extent of the problem, and provided toolkits to providers around how to lower their rates.[1] It is widely accepted that the risk of developing a pressure injury can be reduced by best practices such as frequent repositioning, proper skin care, and specialized cushions or beds; [2,3] studies have also begun to assess the impact of nutritional interventions.[4]

AHRQ published data that showed 3.1 million fewer incidents of hospital-acquired harm in 2011-2015 compared with 2010; 23% of this reduction was from a reduction in-hospital acquired pressure injuries.[1] A 3-year, intervention study found that implementation of a novel 7-step care-based process, acquisition of specialized equipment, and educational initiatives were associated with a significant decrease in incidence rate of pressure injuries.[5] A second study also showed a link between a hospital's processes of care and the outcome of hospital-acquired pressure injury. Processes of care analyzed included risk/skin assessment, risk status at admission, and pressure injury prevention strategies (such as pressure relief). [3]

Early identification and effective facility-level prevention strategies are essential in health care systems for patients at risk for pressure injuries.[6] Further, studies suggest that variation in care delivered negatively impacts pressure injury rates.[7,8] Although the National Pressure Ulcer Advisory Panel (NPUAP) Board of Directors revised the pressure injury staging system in 2015, inaccurate staging of pressure injuries persists impacting the hospital care delivered to patients and influencing their pressure injury rates.[9]

References:

1. Agency for Healthcare Research and Quality. National Scorecard on Rates of Hospital-Acquired Conditions 2010 to 2015: Interim Data from National Efforts to Make Health Care Safer. 2016; https://www.ahrq.gov/professionals/quality-patient-safety/pfp/2015-interim.html?utm_source=AHRQ&utm_medium=PSLS&utm_term=&utm_content=14&utm_campaign=AHRQ_NSOHAC_2016. Accessed January 13, 2017.
2. Agency for Healthcare Research and Quality. Preventing Pressure Ulcers in Hospitals - A Toolkit for Improving Quality of Care. 2012; Rockville, MD. http://www.ahrq.gov/professionals/systems/hospital/pressureulcertoolkit/index.html
3. Gunningberg L, Donaldson, N., Aydin, C., Idvall, E. Exploring variation in pressure ulcer prevalence in Sweden and the USA: Benchmarking in action. 18. 10.1111/j.1365-2753.2011.01702.x. Journal of evaluation in clinical practice. 2011: 904-910.
4. Lozano-Montoya I, Vélez-Díaz-Pallarés M, Abraha I, et al. Nonpharmacologic Interventions to Prevent Pressure Ulcers in Older Patients: An Overview of Systematic Reviews (The Software Engine for the Assessment and optimization of drug and non-drug Therapy in Older persons [SENATOR] Definition of Optimal Evidence-Based Non-drug Therapies in Older People [ONTOP] Series). *J Am Med Dir Assoc.* 2016;17(4):370.e371-370.e310.
5. Lam C, Elkbuli A, Benson B, et al. Implementing a Novel Guideline to Prevent Hospital-Acquired Pressure Ulcers in a Trauma Population: A Patient-Safety Approach. J Am Coll Surg. 2018;226(6):1122-1127.
6. Lyder CH, Ayello EA. Pressure Ulcers: A Patient Safety Issue. In: Hughes RG, editor. Patient Safety and Quality: An Evidence-Based Handbook for Nurses. Rockville (MD): Agency for Healthcare Research and Quality (US); 2008 Apr. Chapter 12. Available from: https://www.ncbi.nlm.nih.gov/books/NBK2650/
7. Edsberg LE, Black JM, Goldberg M, McNichol L, Moore L, Sieggreen M. Revised National Pressure Ulcer Advisory Panel Pressure Injury Staging System: Revised Pressure Injury Staging System. *J Wound Ostomy Continence Nurs.* 2016;43(6):585-597.
8. Horn S, Buerhaus P, Bergstrom N, J. Smout R. RN Staffing Time and Outcomes of Long-Stay Nursing Home Residents. Am J Nurs. 2005; 105:58–70.

9. National Pressure Ulcer Advisory Panel and European Pressure Ulcer Advisory Panel. Prevention and Treatment of Pressure Ulcers: Clinical Practice Guideline. Washington, DC: National Pressure Ulcer Advisory Panel. 2014.

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

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

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	
Provide all other grades and definitions from the evidence grading system	
Grade assigned to the **recommendation** with definition of the grade	

Provide all other grades and definitions from the recommendation grading system	
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	
Estimates of benefit and consistency across studies	
What harms were identified?	
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	

1a.4 OTHER SOURCE OF EVIDENCE

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

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

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

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

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- Disparities in care across population groups.

1b.1. Briefly explain the rationale for this measure *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*

If a COMPOSITE (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and answer the composite questions.

This safety eCQM captures the number of patients who experience harm in the form of a pressure injury, during their inpatient hospitalization. Hospital-acquired pressure injuries are serious events and one of the most common patient harms. Pressure injuries commonly cause local infection, osteomyelitis, anemia, and sepsis (Brem, et al., 2010), in addition to causing significant depression, pain, and

discomfort to patients (Gunningberg et al., 2011). Pressure injury is considered a serious reportable event by the National Quality Forum (NQF) (Centers for Medicare and Medicaid Services, 2015). CMS also established non-payment for pressure injury (National Quality Forum, 2016), and the rate of pressure injuries is considered an indicator of the quality of nursing care a hospital provides (National Quality Forum, 2005).

It is widely accepted that the risk of developing a pressure injury can be reduced through best practices such as frequent repositioning, proper skin care, and specialized cushions or beds (Berlowitz, et al., 2012). Systematically measuring patients who develop new pressure injuries while in the hospital setting will provide hospitals with a reliable and timely measurement, to more reliably assess harm reduction efforts and modify their improvement efforts in near real-time. This eCQM will fill a gap in measurement and provide incentives for hospitals' quality improvement. Although several pressure injury measures are currently in use, there are no electronic health record (EHR)-based measures intended for use in acute care hospitals. In addition, the intent of this measure is to incentivize greater achievements in reducing harms and enhance hospital performance on patient safety outcomes.

References:

Brem H, M. J., Nierman D, et al. (2010). High Cost of Stage IV Pressure Ulcers. doi:10.1016/j.amjsurg.2009.12.021. American Journal of Surgery, 200(4), 473-477.

Gunningberg, L., Donaldson, N., Aydin, C., Idvall, E. (2011). Exploring variation in pressure ulcer prevalence in Sweden and the USA: Benchmarking in action. 18. 10.1111/j.1365-2753.2011.01702.x. Journal of evaluation in clinical practice., 904-910.

Centers for Medicare & Medicaid Services. (2015). Hospital-Acquired Conditions. Retrieved January 13, 2017, from https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HospitalAcqCond/Hospital-Acquired_Conditions.html

National Quality Forum. (2016). List of SREs. Retrieved January 13, 2017, from http://www.qualityforum.org/Topics/SREs/List_of_SREs.aspx#sre4

National Quality Forum. (2005). National Voluntary Consensus Standards for Nursing-Sensitive Care: An Initial Performance Measure Set. Retrieved January 13, 2017, from http://www.qualityforum.org/Publications/2004/10/National_Voluntary_Consensus_Standards_for_Nursing-Sensitive_Care__An_Initial_Performance_Measure_Set.aspx.

Berlowitz, D. VanDeusen Lukas, C.; Parker, V.; Niederhauser, A.;, & Silver, J. L., C.; Ayello, E.; Zulkowski, K. (2012). Preventing Pressure Ulcers in Hospitals- A Toolkit for Improving Quality of Care.

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

This eCQM was tested with 3 test sites (24 hospitals) in 3 states (located in Midwest, West, and Northeast). Hospitals varied in size (200+ beds, 15-500 beds, and 450-700 beds), EHR systems (Meditech, Cerner, Epic), teaching status (teaching and non-teaching hospitals), and location (urban, suburban, and rural). A detailed breakdown of the characteristics of the measured facilities and the patient population can be found in the attached Measure Testing Form (Beta Datasets 1, 2, and 3).

The measure performance, including the denominator, numerator, and measure rate by hospital, follows.

Hospital Test Site 1 (Beta Dataset 1 per Testing Form)

- Number of Hospitals: 1
- Data collection period: 1/1/2017 - 12/31/2017
- Denominator: 7,573
- Numerator: 38
- Performance rate: 0.50%
- 95% confidence interval: 0.36%, 0.69%
- Standard Deviation: N/A (only one hospital)

Hospital Test Site 2 (Beta Dataset 2 per Testing Form)

- Number of Hospitals: 21
- Data collection period: 1/1/2017 - 12/31/2017
- Denominator: 100,238
- Numerator: 724
- Performance rate: 0.72%
- 95% confidence interval: 0.67%, 0.78%
- Standard Deviation: 0.47%

Hospital Test 3 (Beta Dataset 3 per Testing Form)

- Number of Hospitals: 2
- Data collection period: 1/1/2017 - 12/31/2017
- Denominator: 56, 330
- Numerator: 414
- Performance rate: 0.73%
- 95% confidence interval: 0.67%, 0.81%
- Standard Deviation: 0.06%

Overall Performance

- Number of Hospitals: 24
- Performance rate: 0.72%
- 95% confidence interval: 0.68%, 0.76%
- Standard deviation: 0.45%
- Range: 0.0% to 1.46%

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.

N/A

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for maintenance of endorsement. Describe the data source including number of*

measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.

Data below are from initial development testing; this eCQM is not yet implemented. The measure performance was stratified for disparities by age, race, ethnicity, and payer source.

Hospital Test Site 1 (Beta Dataset 1 per Testing Form)

- Number of hospitals: 1
- Data collection period: 1/1/2017 - 12/31/2017
- Denominator (admissions): 7,573

Hospital Test Site 2 (Beta Dataset 2 per Testing Form)

- Number of hospitals: 21
- Data collection period: 1/1/2017 - 12/31/2017
- Denominator (admissions): 100,238

Hospital Test Site 3 (Beta Dataset 3 per Testing Form)

- Number of hospitals: 2
- Data collection period: 1/1/2017 - 12/31/2017
- Denominator (admissions): 56,330

Category//Denominator//Numerator//Measure Rate (95% Confidence Interval)

Across Sites (n=164,141, 24 hospitals)

Age//Denominator//Numerator//Measure Rate (95% Confidence Interval)

18-64//104,332//401//0.38% (0.3%, 0.4%)

65+//59,809//775//1.30% (1.2%, 1.4%)

Gender//Denominator//Numerator//Measure Rate (95% Confidence Interval)

Male//61,636//664//1.08% (1.0%, 1.2%)

Female//102,503//512//0.50% (0.5%, 0.5%)

Unknown//2//0//0.00% (0.0%, 0.7%)

Race//Denominator//Numerator// Measure Rate (95% Confidence Interval)

Black or African American//7,195//51//0.71% (0.5%, 0.9%)

White//133,894//974//0.73% (0.7%, 0.8%)

Other//21,795//142//0.65% (0.5%, 0.8%)

Unknown//1,257//9//0.72% (0.3%, 1.4%)

Ethnicity//Denominator//Numerator//Measure Rate (95% Confidence Interval)

Hispanic or Latino//18,030//89//0.49% (0.4%, 0.6%)

Non-Hispanic//142,251//1,057//0.74% (0.7%, 0.8%)

Unknown//3,860//30//0.78% (0.5%, 1.1%)

(Primary) Payer//Denominator//Numerator// Measure Rate (95% Confidence Interval)

Medicare//64,913//806//1.24% (1.2%, 1.3%)

Medicaid//12,280//96//0.78% (0.6%, 1.0%)

Private Insurance//75,895//236//0.31% (0.3%, 0.4%)

Self-pay or Uninsured//5,999//9//0.15% (0.1%, 0.3%)

Other (such as other government plans)//4,475//27//0.60% (0.4%, 0.9%)

Unknown//579//2//0.35% (0.0%, 1.2%)

It is important to note these results are derived from a small dataset that is not generalizable to the entire population, and the datasets include many characteristics that are 'unknown' in the EHR, which limits the usability of the results.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b.4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in 1b.4

N/A

2. Reliability and Validity—Scientific Acceptability of Measure Properties

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

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

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

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

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

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

Final measure specifications for implementation will be made publicly available on CMS' appropriate quality website, once finalized through the NQF endorsement and CMS rulemaking processes.

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:** Del18c3HOP5HarmsPressureInjuryITS12172018_v5_6_Artifacts-636824580499913599.zip,Pressure_Injury_Bonnie_Test_Cases_Results.pdf

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:** Del18c3HOP5HarmsPressureInjuryFeasibilityScorecard12172018_v02-636824582773862567.xlsx

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

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

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

Not an instrument-based measure

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

No

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

N/A

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

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

The number of hospital inpatient admissions during which a patient developed a new stage 2, stage 3, stage 4 pressure injury, deep tissue pressure injury, or unstageable pressure injury that was not documented as present in the first 24 hours of hospital arrival.

S.5. Numerator Details *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*

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

This is an eCQM, and therefore uses electronic health record data to calculate the measure score. The time period for data collection is during an inpatient hospitalization, beginning at hospital arrival (whether through Emergency Department, observation stay, or directly admitted as inpatient). All data elements necessary to calculate this measure are defined within value sets, described below and available in the VSAC.

Pressure ulcer stage is defined by the VSAC as Pressure Ulcer Stage (2.16.840.1.113883.11.20.9.35).

To access the value sets for the measure, please visit the Value Set Authority Center (VSAC), sponsored by the National Library of Medicine, at https://vsac.nlm.nih.gov/.

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

All patients 18 years or older at the start of the encounter and discharged inpatient hospital admission during the measurement period. The measure includes inpatient admissions which began in the Emergency Department or in observational status.

S.7. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection*

items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)

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

This measure includes all inpatient admissions for patients aged 18 years and older at the time of admission, and all payers. Measurement period is one year. This measure is at the hospital-by-admission-level; only one numerator event is counted per admission.

Inpatient Encounters are represented using the value set of Encounter Inpatient (2.16.840.1.113883.3.666.5.307).

Emergency Department visits are represented using the value set of Emergency Department Visit (2.16.840.1.113883.3.117.1.7.1.292).

Patients whom had observation encounters are represented using the value set of Observation Services (2.16.840.1.113762.1.4.1111.143).

To access the value sets for the measure, please visit the Value Set Authority Center (VSAC), sponsored by the National Library of Medicine, at https://vsac.nlm.nih.gov/.

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

There are no denominator exclusions.

S.9. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

N/A; there are no denominator exclusions.

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

N/A; this measure is not stratified.

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

No risk adjustment or risk stratification

If other:

S.12. Type of score:

Rate/proportion

If other:

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

Better quality = Lower score

S.14. Calculation Algorithm/Measure Logic (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the*

target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)

Target population

Inpatient admission encounters, all payer, where individuals are aged 18 years or older at the start of the admission and are discharged within the measurement period.

To create the denominator:

1. If the inpatient admission was during the measurement period, go to Step 2. If not, do not include in measure population.

2. Determine the patient's age in years. The patient's age is equal to the admission date minus the birth date. If the patient is 18 years or older, include in the measure population. If less than 18 years old, do not include in the measure population.

To create the numerator:

1. Of encounters in the denominator, include any qualifying inpatient admissions which include a stage 2, stage 3, stage 4, deep tissue pressure injury, or unstageable pressure injury that was not documented within first 24 hours after hospital arrival.

2. Of the events, keep one (the first) qualifying event per encounter. This measure counts one harm per encounter.

See algorithm flowchart attached as appendix.

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

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

N/A; this measure does not use a sample.

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

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

N/A; this measure does not use a survey.

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

If other, please describe in S.18.

Electronic Health Records

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

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

Hospitals collect EHR data using certified electronic health record technology (CEHRT). The MAT output, which includes the human readable and XML artifacts of the clinical quality language (CQL) for the measure are contained in the eCQM specifications attached. No additional tools are used for data collection for eCQMs.

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

No data collection instrument provided

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

Facility

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

Inpatient/Hospital

If other:

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

N/A

2. Validity – See attached Measure Testing Submission Form

Del18c3HOP5HarmsPressureInjuryITSForm010219.docx,Del18c3HOP5HarmsPressureInjuryTestingForm012219_v1.0.docx

2.1 For maintenance of endorsement

Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

2.2 For maintenance of endorsement

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

2.3 For maintenance of endorsement

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

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

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b1-2b6)

Measure Number (*if previously endorsed*)**:** N/A
Measure Title: Hospital Harm- Pressure Injury
Date of Submission: TBD
Type of Measure:

☒ Outcome (*including PRO-PM*)	☐ Composite – ***STOP – use composite testing form***

☐ Intermediate Clinical Outcome	☐ Cost/resource
☐ Process *(including Appropriate Use)*	☐ 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, 2b1, 2b2, and 2b4 must be completed.**
- **For outcome and resource use measures**, section **2b3** also must be completed.
- If specified for **multiple data sources/sets of specifications** (e.g., claims and EHRs), section **2b5** 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 (2b1-2b6) 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 25 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 social risk factors variables and testing in this form refer to the release notes for version 7.1 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Standing 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 **instrument-based measures** (including PRO-PMs) **and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b1. 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 **instrument-based measures (including PRO-PMs) and composite performance measures**, validity should be demonstrated for the computed performance score.

2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure; [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]

2b3. 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 social risk 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.

2b4. 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.

2b5. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b6. 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. The degree of consensus and any areas of disagreement must be provided/discussed.

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

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

We acquired data from a patient safety organization to support alpha testing of the measure concept, data elements, and validity. We partnered with two health systems to complete beta testing of the MAT output in three different EHR systems. We assessed data element and measure score validity as well as measure score reliability in beta testing. The dataset used varies by testing type; see Section 1.7 for details.

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

The dates vary by testing type; see Section 1.7 for details.

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

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

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

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

The number of measured entities (hospitals) varies; see Section 1.7 for details.

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*)

The number of admissions/patients varies; see Section 1.7 for details.

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

The datasets, dates, number of measured hospitals, and number of admissions used in each phase of testing are in **Table 1**.

Table 1. Dataset Descriptions

Dataset	Applicable Section in the Testing Attachment	Description of Dataset	EHR Vendor
Beta Dataset 1	Section 2a2 Reliability Testing Section 2b1 Validity Testing Section 2b4 Identification of Statistically Significant and Meaningful Differences in Performance Section 2b6 Missing Data Analysis	Dates of Data: January 1, 2017 - December 31, 2017 Number of Hospitals: 1 Number of Admissions: 7,573 Number of Unique Patients: 5,735 For Validity Testing: sample of 186 admissions Hospital was in suburban location, with over 200 beds, and not a teaching hospital. Located in the Midwest.	Meditech

Dataset	Applicable Section in the Testing Attachment	Description of Dataset	EHR Vendor
Beta Dataset 2	Section 2a2 Reliability Testing Section 2b1 Validity Testing Section 2b4 Identification of Statistically Significant and Meaningful Differences in Performance Section 2b6 Missing Data Analysis	Dates of Data: January 1, 2017 - December 31, 2017 Number of Hospitals: 21 Number of Admissions: 100,238 Number of Unique Patients: 84,745 For Validity Testing: sample of 175 admissions Hospitals were within one health system, mixed between urban, suburban, and rural locations, some being teaching hospitals. Hospitals ranged between 15 – 500 beds. Located in the West.	Cerner
Beta Dataset 3	Section 2a2 Reliability Testing Section 2b1 Validity Testing Section 2b4 Identification of Statistically Significant and Meaningful Differences in Performance Section 2b6 Missing Data Analysis	Dates of Data: January 1, 2017 - December 31, 2017 Number of Hospitals: 2 Number of Admissions: 56,330 Number of Unique Patients: 45,699 For Validity Testing: sample of 173 admissions Hospitals were within one health system, in urban location, some teaching hospitals. Hospitals	Epic

Dataset	Applicable Section in the Testing Attachment	Description of Dataset	EHR Vendor
		ranged between 450 – 700 beds. Located in the Northeast.	
Alpha Dataset	Section 2b1 Validity Testing (Measure Score)	Dates of Data: June 1, 2016 - May 31, 2017 Number of hospitals: 5 Number of Admissions: 66,127 Hospitals were in two different health systems, both in urban locations, and not-for-profit. They were diverse in terms of bed size (between 100-199 beds and 300-399 bed), teaching status, and geographic location (South, West).	Cerner & Epic

Patient descriptive characteristics in Alpha Dataset are as follows:

- Patient Descriptive Characteristics:
  - Mean age at admission = 58.7 years with a standard deviation of 20.4 years
  - 58.2% female, 41.8% male
  - 64.5% White, 9.7% Black or African American, 8.0% Asian, 1.0% Native Hawaiian or Other Pacific Islander, 0.2% American Indian or Alaska Native, 15.7% Other, and 0.9% declined or unknown

Patient descriptive characteristics included in the analysis by hospital for **Beta Datasets 1, 2, and 3** are provided in **Table 2.**

Table 2. Demographic Characteristics of Eligible Patient Population (Beta Datasets 1, 2, and 3)

Initial Patient Population Characteristics	Beta Dataset 1 (N, %)	Beta Dataset 2 (N, %)	Beta Dataset 3 (N, %)	Across Beta Sites (N, %)
Number of unique patients	5,735, 100%	84,745, 100%	45,699, 100%	136,179, 100%
Average Age [Mean (STD)]	56 (22)	50 (22)	54 (21)	52 (21)

Initial Patient Population Characteristics	Beta Dataset 1 (N, %)	Beta Dataset 2 (N, %)	Beta Dataset 3 (N, %)	Across Beta Sites (N, %)
18-35	1,455, 25.4%	31,386, 37.0%	12,115, 26.5%	44,956, 33.0%
36-64	1,944, 33.9%	26,365, 31.1%	16,841, 36.9%	45,150, 33.2%
65+	2,336, 40.7%	26,994, 31.9%	16,743, 36.6%	46,073, 33.8%
Sex				
Male	2,212, 38.6%	28,694, 33.9%	17,213, 37.7%	48,119, 35.3%
Female	3,523, 61.4%	56,050, 66.1%	28,485, 62.3%	88,058, 64.7%
Unknown	0, 0.0%	1, 0.0%	1, 0.0%	2, 0.0%
Race				
Black or African American	146, 2.6%	825, 1.0%	4,849, 10.6%	5,820, 4.3%
White	5,540, 96.6%	79,042, 93.3%	26,608, 58.2%	111,190, 81.7%
Other	26, 0.5%	4,871, 5.8%	13,167, 28.8%	18,064, 13.3%
Unknown	23, 0.4%	7, 0.0%	1,075, 2.4%	1,105, 0.8%
Ethnicity				
Hispanic or Latino	32, 0.6%	7,880, 9.3%	7,080, 15.5%	14,992, 11.0%
Non-Hispanic	5,589, 97.5%	75,575, 89.2%	36,513, 79.9%	117,677, 86.4%
Unknown	114, 2.0%	1,290, 1.5%	2,106, 4.6%	3,510, 2.6%
(Primary) Payer				
Medicare	2,699, 47.1%	35,107, 41.4%	12,099, 26.5%	49,905, 36.7%
Medicaid	1,490, 26.0%	3,931, 4.6%	4,576, 10.0%	9,997, 7.3%
Private Insurance	1,336, 23.3%	38,396, 45.3%	27,389, 59.9%	67,121, 49.3%
Self-pay or Uninsured	188, 3.3%	4,773, 5.6%	0, 0.0%	4,961, 3.6%
Other+	22, 0.4%	2,538, 3.0%	1,111 2.4%	3,671, 2.7%
Unknown	0, 0.0%	0, 0.0%	524, 1.2%	524, 0.4%

+ “Others” include all possible payers other than Medicare and Medicaid, such as other government plans (e.g. federal, state, local), private health insurance, etc.

1.8 What were the social risk factors that were available and analyzed? For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.

As described in Section 1.7, **Table 1**, we collected information on the following social risk factors using data extracted from hospital EHR systems: race, ethnicity, and primary payer.

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*)

Data Element Reliability

N/A. Since data element validity was empirically tested, separate reliability testing of data elements is not required per the NQF Measure Evaluation Criteria and Guidance (see section 2b2 for validity testing of data elements).

Measure Score Reliability

The reliability of a measure score is the degree to which repeated measurements of the same entity agree with each other. We estimated the measure score reliability using **Beta Datasets 1, 2, and 3**. We assessed signal-to-noise reliability using **Beta Datasets 1, 2, and 3** that describes how well the measure can distinguish the performance of one hospital from another (Adams and Mehrota, 2010; Yu and Mehrota, 2013). The signal is the proportion of the variability in measured performance that can be explained by real differences in performance. Scores can range from 0 to 1. 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.

We use the Adam's beta-binomial method (Adams, 2009) to calculate the signal-to-noise ratio reliability. Briefly, using variability between hospitals (signal: provider-to-provider variance) and variability within hospitals (noise: provider-specific-error variance), the reliability for each hospital can be defined as

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

We estimate the beta-binomial variance as the provider-to-provider variance as

$$\sigma^2_{provider-to-provider} = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$$

where α, β are the estimated beta-binomial parameters using denominators and rates from all hospitals. The provider-specific-error variance is estimated as

$$\sigma^2_{provider-specific-error} = \frac{\hat{p}(1-\hat{p})}{n}$$

where n is the numerator of a hospital and pˆ is the harm rate of a hospital.

References:

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

Yu, H, Mehrota, A, Adams J. (2013). Reliability of utilization measures for primary care physician profiling. Healthcare, 1, 22-29.

Adams, J. The Reliability of Provider Profiling: A Tutorial. Santa Monica, CA: RAND Corporation, 2009. https://www.rand.org/pubs/technical_reports/TR653.html.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Measure Score Reliability
There were 164,141 eligible encounters across 24 hospitals in **Beta Datasets 1, 2, and 3**. The signal-to-noise ratio yielded a median reliability score of 0.969 (range: 0.206 to 1.000).

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 signal-to-noise ratio of 0.969 indicates excellent agreement.

Our interpretation of these results is based on the standards established by Landis and Koch (1977):
< 0 – Less than chance agreement;
0 – 0.2 Slight agreement;
0.21 – 0.39 Fair agreement;
0.4 – 0.59 Moderate agreement;
0.6 – 0.79 Substantial agreement;
0.8 – 0.99 Almost Perfect agreement; and
1 Perfect agreement

Reference:
Landis J, Koch G. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-174.

2b1. VALIDITY TESTING

2b1.1. What level of validity testing was conducted? (*may be one or both levels*)

☒ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☒ **Empirical validity testing**

☐ **Systematic assessment of face validity of <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*) **NOTE**: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

2b1.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

Data element validity was assessed by evaluating the accuracy of electronically extracted EHR data elements compared with manually chart abstracted data elements for the same patients, which is considered the "gold standard" for the purpose of these analyses.

<u>Data Element Validity:</u>

For **Beta Datasets 1, 2, and 3**, a stratified random sample of total admissions were selected at each hospital test site. Sample size calculations ensure a robust sample was used for validity testing. Specifically, we derived our sample size based on the following assumptions: Our primary endpoint for sample size estimation is PPV, which is applicable for both data element validity and measure score validity. We adjudicated all our numerator cases in alpha test and obtained high PPVs (>90% in most of the cases). Based on this, we approximate the sample size based on one-sample proportion formula as the following:

$n=(moe/z_{(\alpha/2)})^2 * p * (1-p)$

Where *a* is the type I error rate, *moe* is the margin of error, p is the proportion, here PPV, of interest. We simulate a series of *moe* and target PPV values for sample size and 95% confidence interval (CI) estimation. For example, with a *moe* of 6% and a target PPV of 0.9, a sample size of 100 will give rise to a 95% CI of 0.84 – 0.96. We concluded that a sample size of 100 from each hospital would ensure an accurate PPV estimation. Also, combining the samples from more than 1 hospitals would give us even more accurate estimation.

Beta Dataset 1 had 186 encounters, 38 being admissions with harm events and 148 being admissions without a harm event (denominator-only); **Beta Dataset 2** had 175 encounters, 100 being admissions with harm events and 75 being admissions without a harm event (denominator-only); and **Beta Dataset 3** had 173 encounters, 98 being admissions with harm events and 75 being admissions without a harm event (denominator-only). Data were abstracted from the EHR by trained abstractors at each test site; abstractors at all sites had experience abstracting data for chart-based quality measure reporting. Abstractors were provided with an instruction manual and an Access database to document the information abstracted from the EHR. Access databases were only pre-populated with the unique patient identifier; abstractors were asked to input all

other data from the chart independently of the EHR dataset. Abstraction training was also provided to each site.

Table 3 shows the PPV agreement rate (# exact matches in both data sources / # in the EHR) between the data extracted from the EHR electronically and manual chart abstraction in **Beta Datasets 1, 2, and 3**. Each data element matched if the manually abstracted value exactly matched the specific electronically extracted value. For example, out of 275 specific instances where a patient had a pressure injury with date/time (in the EHR data), 252 of those specific pressure injuries exactly matched with data and time in the abstracted chart data (gold standard), resulting in an 91.6% match rate. For data/time data elements, we matched month, day, year, hour, and minutes.

Empirical Measure Score Validity:
Measure score validity assesses whether the harm rate (or, the measure score outcome) calculated for each facility is in fact accurate. The measure score is calculated for each facility based on the number of encounters that experienced a harm compared to the total number of encounters. Therefore, we validated each individual harm identified in a sample of cases in the EHR by chart review by trained abstractors to confirm that the chart, or gold standard, reflects that a harm occurred. Because no further calculations are conducted to generate a facility level score (as is with risk-adjusted measures), We did not compare the harm rate to any other external measure of quality. For measures that count harm events without other statistical manipulation, the confirmation that the measure logic is accurately capturing true harm events is the gold standard for assessing validity of the measure score.

Therefore, to validate the EHR-extracted numerator against the gold standard of the patient medical chart, to assess whether the harms actually occurred and captured the intended outcome, we clinically adjudicated each admission that met the criteria for a harm among the sample of abstracted records, and calculated the positive predictive value (PPV) for all numerator cases, and denominator cases, as shown in **Table 5,** in **Alpha Dataset, Beta Datasets 1, 2 and 3**. The PPV describes the probability that a patient with a positive result (numerator case) in the EHR data also was a positive result in the abstracted medical record data, as confirmed by a clinical adjudicator. Similarly, for denominator cases, the PPV describes the probability that a patient was identified as a denominator case in the HER was also a denominator case in the chart abstracted medical record data.

We also calculated the sensitivity, specificity, kappa, and negative predictive value (NPV) as shown in **Table 4** for Beta Dataset 1, 2 and 3. Sensitivity describes the probability that a patient with a positive result in the abstracted medical record data was also a positive result in the EHR data. Specificity describes the probability that a patient with a negative result (not a numerator case) in the abstracted medical record data was also a negative result in the EHR data. Kappa describes the amount of remaining agreement between the harm incidences based on EHR and the harm incidences based on the abstracted medical record after the agreement by chance is taken into account. NPV describes the probability that a patient with a negative result (not in the numerator) in the EHR data also was a negative result in the abstracted medical record, confirmed by the clinical adjudicator.

For **Alpha Dataset**, data were abstracted from the EHR by trained abstractors who had experience abstracting data for chart-based quality measure reporting. Abstractors were provided with an instruction manual and an Excel, to document the information abstracted from the EHR. Abstraction training was also provided. Validity was established in **Beta Datasets 1, 2** and **3** as described above.

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

Data Element Validity

Table 3. Data Element Validity (PPV) Results Required for Measure (Beta Dataset 1, 2, and 3)

Data Element	Beta Dataset 1			Beta Dataset 2			Beta Dataset 3		
	# Cases Matched in Abstraction (n)	# Cases Matched in EHR (n)	PPV Percent Match (%)	# Cases Matched in Abstraction (n)	# Cases Matched in EHR (n) (n)	Percent Match (%)	# Cases Matched in Abstraction (n)	# Cases Matched in EHR (n)	PPV Percent Match (%)
Admission date and time (mm/dd/yyyy hh:mm)	291	291	100.0%	275	275	100.0%	272	272	100%
Diagnosis: Pressure injury date and time	387	533	72.6%	252	275	91.6%	257	272	94.5%
Diagnosis: Pressure injury stage	459	533	86.1%	251	275	91.3%	260	272	95.6%
Patient characteristic: birth date	283	291	97.3%	272	275	98.9%	267	272	98.2%

Empirical Measure Score Validity

Table 4 displays the specificity, sensitivity, kappa, and NPV in each Beta Dataset. **Table 5** displays the positive predictive value (PPV) in each dataset. This PPV represents the percent of admissions that met the criteria for a harm (numerator)in the EHR confirmed by the chart abstraction, validated by a trained clinical adjudicator. **Alpha Dataset** validated the numerator cases and not denominator cases, due to data limitations. **Beta Datasets 1, 2 and 3** were able to validate both numerator and denominator.

Table 4. Measure Score Validity Statistics for Sample Between Electronic EHR Extraction and Manual Chart Abstraction (Sensitivity, Specificity, NPV, Kappa) (Bata Datasets 1, 2, 3)

Measure	Beta Dataset 1	Beta Dataset 2	Beta Dataset 3

	Sensitivity	Specificity	Kapa (95% CI)	NPV	Sensitivity	Specificity	Kapa (95% CI)	NPV	Sensitivity	Specificity	Ka
Pressure Injury	84%	92%	0.70 (0.56, 0.83)	97%	98%	96%	0.94 (0.89, 0.99)	97%	100%	58%	

Table 5. Measure Score Validity Statistics for Sample Between Electronic EHR Extraction and Manual Chart Abstraction (PPV) (Alpha Dataset, Beta Datasets 1, 2. And 3)

Measure Component	Alpha Dataset PPV	Beta Dataset 1 PPV	Beta Dataset 2 PPV	Beta Dataset PPV 3
Initial patient population/denominator	N/A	100%	100%	100%
Numerator	97.8%	68.4%	97.0%	44.9%

2b1.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

Data Element Validity

All but one data element had a match rate of 86% and higher, with most over 91.3%, indicating valid and accurate data elements were extracted from the EHR. The exception was at **Beta Dataset 1**, pressure injury with date and time, of 72.6%. We believe this specific match rate was due discrepancies by the abstractor, whether they noted the original pressure injury date and time, or when that pressure injury was staged; additionally, many pressure injuries at this hospital were originally documented in notes fields, contributing to a third option for date and time in abstraction. This rationale also applies to the pressure injury stage, as it was often documented in clinical notes, which is not an extractable field but could be reviewed by abstractors. Overall, we believe the data elements required for the measure show validity.

Empirical Measure Score Validity

In **Alpha Dataset** and **Beta Dataset 2**, PPV was high at 97.8% and 97%, respectively, meaning that in almost all cases, the admissions met the criteria for a harm in the chart abstracted and EHR data. **Beta Dataset 1** and **Beta Dataset 3** had lower PPV at 68.4% and 44.9% respectively. Understanding the reasons for the discrepancy lends to the fuller picture, which are as follows:

- Majority of discrepancies (6 out of 12 cases in Beta Dataset 1; 53 out of 54 cases in Beta Dataset 3) were documented in free text (clinical notes) within the first 24 hours of admission, and therefore were not captured by the EHR because these were not in structured fields;
- A couple were due to abstraction error, where the date of the pressure injury was not provided;
- One was due to an EHR extraction error, this case was a skin tear and not a pressure injury;
- A few encounters had pressure injuries in the EHR, but the abstractor did not find the information in the chart.

The primary source of lower PPV in 2 of the 3 health systems was documentation practices, where pressure injuries were documented only in text notes and not in structured fields within the first 24 hours of hospital arrival. Because the injury was documented in a structured field only after 24 hours had elapsed, the measure logic identified these pressure injuries as new. The measure logic therefore worked as intended. However, hospitals' failure to use structured fields for initial documentation effected their harm rate. That pattern of documentation was counted as a harm in the measure rate. **If we remove these cases from the PPV calculation, PPV is 81.3% and 97.8% in Beta Dataset 1 and 3 respectively. This indicates that the measure is highly valid when hospitals consistently use structured fields to document pressure injuries.**

Although we do not always expect perfect agreement, as we expect some degree of human error in entering and matching values, our results suggest that the measure is valid and provides an indication of quality of care provided by the hospital. The absence of a perfect PPV does not threaten validity as we do not expect any systematic error in this disagreement across hospitals that might bias the measure results. We will continue to reevaluate validity through reevaluation as hospitals participate in this measure and as required by NQF for maintenance of endorsement.

In Beta Datasets 1, 2 and 3 sensitivity is high at 84-98%. This means that the probability of the EHR data detecting a new pressure injury in patients that had a true new pressure injury based on the abstracted data ('gold standard') ranges from 84% to 98% (sensitivity). The probability of the EHR data detecting no new pressure injury out of no new pressure injury patients based on abstracted data is (specificity) is high in Beta Dataset 1 and 2 (92% and 96%, respectively). The low specificity is due to high number of false positive cases caused by documentation error (documented in unstructured fields, clinical notes in the EHR) in Beta Dataset 3. NPV was 97%-100% in all three Beta Datasets, indicating the EHR data indicated a harm did not occur, and almost all of the time the chart abstraction confirmed a harm did not occur. Kappa of 0.70 and 0.94 in Beta Dataset 1 and 2 indicates substantial to excellent agreement. Kappa was lower in Beta Dataset 3.

Our Kappa interpretation is based on the following standards set by Viera et al.:
0.4 – 0.6 indicate "moderate agreement",
0.6 – 0.8 "substantial agreement", and
0.8 – 1 "almost perfect agreement"

References:
1. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960;20:37–46.
2. Viera AJ, Garrett JM. Understanding Interobserver Agreement: The Kappa Statistic. Fam Med 2005;37(5):360-3.

2b2. EXCLUSIONS ANALYSIS

NA ☒ no exclusions — *skip to section* 2b3

2b2.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

2b2.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

2b2.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

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

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

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

☒ **No risk adjustment or stratification**

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

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

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

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

2b3.2. If an outcome or resource use component measure is <u>not risk adjusted or stratified</u>, provide <u>rationale and analyses</u> to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

Clinical characteristics, including a patient's age, reason for hospitalization, clinical status when they arrive at the hospital, or comorbid conditions all may influence the risk of harm occurring during a hospitalization. Therefore, if hospitals care for patients with different degree of risk, it may be important to adjust for patient risk factors in order to compare hospital performance. However, many harms should be avoidable, regardless of patient risk. We consider the following criteria in determining whether risk adjustment is warranted:

1. If many patients are at risk of the harm regardless of their age, clinical status, comorbidities, or reason for admission.
2. If the majority of incidents of the harm are linkable to care provision under the control of providers, for example harms caused by excessive or inappropriate medication dosing.
3. If there is evidence that the risk of a harm can be largely ameliorated by best care practices regardless of a patients' inherent risk profile. For example, there may be evidence that even complex patients with multiple risk factors can avoid harm events when providers closely adhere to care guidelines.

In the case of this hospital-acquired pressure injury eCQM, there is evidence indicating that most newly acquired pressure injuries are avoidable with best practice. Although certain patients may be particularly vulnerable to pressure injuries in certain settings (e.g. permanent or prolonged immobility), the most common causes are limited mobility during an acute illness and friction or shear against sensitive skin, which many hospitalized patients are at risk of these injuries. There are many actions hospitals can take to reduce patient harm risk, such as conducting a structured risk assessment to identify individuals at risk for pressure injury as soon as possible upon arrival and repeating at regular intervals, as well as proper skin care, nutrition, and careful repositioning of patients. As many of the causes can be mitigated through best care in hospital environments, we do not think risk adjustment is warranted for this measure. We will continue to evaluate the appropriateness of risk adjustment in measure reevaluation.

In addition to the clinical rationale provided for not risk adjusting this measure, we examined the performance (harm) rate of the measure across patient characteristics of age, sex, race, ethnicity, and payer. Age (by date of birth) was validated; no other patient demographic was validated using chart data. It is important to note these results are derived from a small dataset that is not generalizable to the entire population and the datasets include many characteristics that are 'unknown' in the EHR which limits the usability of the results; additionally, we do not believe it is clinically appropriate to adjust by these characteristics given the clinical rationale provided above.

Table 6. Performance Rate by Encounter Characteristic (Beta Datasets 1, 2 and 3)

Characteristic	Beta Dataset 1			Beta Dataset 2			Beta Dataset 3		
	Denominator	Numerator	Performance Rate % (95% CI)	Denominator	Numerator	Performance Rate % (95% CI)	Denominator	Numerator	Perform Rat (95
Number of unique Encounters	7,573	38	0.5 (0.4, 0.7)	100,238	724	0.7 (0.7, 0.8)	56,330	414	0.7 (0.7,
Average Age									
18-64	4,193	12	0.3 (0.1, 0.5)	66,331	288	0.4 (0.4, 0.5)	33,808	101	0.3 (0.2,
65+	3,380	26	0.8 (0.5, 1.1)	33,907	436	1.3 (1.2, 1.4)	22,522	313	1.4 (1.2,
Sex									
Male	3,012	22	0.7 (0.5, 1.1)	36,019	409	1.1 (1.0, 1.3)	22,605	233	1.0 (0.9,
Female	4,561	16	0.4 (0.2, 0.6)	64,218	315	0.5 (0.4, 0.5)	33,724	181	0.5 (0.5,
Unknown	0	0	N/A	1	0	0.0 (0.0, 1.0)	1	0	0.0 (0.0, 97.5
Race									
Black or African-American	184	2	1.1 (0.1, 3.9)	1,010	9	0.9 (0.4, 1.7)	6,001	40	0.7 (0.5,
White	7,327	36	0.5 (0.3, 0.7)	93,731	670	0.7 (0.7, 0.8)	32,836	268	0.8 (0.7,
Other	34	0	0.0 (0.0, 10.3)	5,490	45	0.8 (0.6, 1.1)	16,271	97	0.6 (0.5,
Unknown	28	0	0.0 (0.0, 12.3)	7	0	0.0 (0.0, 0.4)	1,222	9	0.7 (0.3,
Ethnicity									
Hispanic or Latino	39	0	0.0 (0.0, 9.0)	9,032	54	0.6 (0.4, 0.8)	8,959	35	0.4 (0.3,
Non-Hispanic	7,405	36	0.5 (0.3, 0.7)	89,818	657	0.7 (0.7, 0.8)	45,028	364	0.8 (0.7,

Characteristic	Beta Dataset 1			Beta Dataset 2			Beta Dataset 3		
	Denominator	Numerator	Performance Rate % (95% CI)	Denominator	Numerator	Performance Rate % (95% CI)	Denominator	Numerator	Perform Rat (95
Unknown	129	2	1.6 (0.2, 5.5)	1,388	13	0.9 (0.5, 1.6)	2,343	15	0.6 (0.4
(Primary) Payer									
Medicare	3,920	29	0.7 (0.5, 1.1)	44,426	552	1.2 (1.1, 1.3)	16,567	225	1.4 (1.2
Medicaid	1,829	5	0.3 (0.1, 0.6)	4,586	40	0.9 (0.6, 1.2)	5,865	51	0.9 (0.6
Private Insurance	1,583	2	0.1 (0.0, 0.5)	42,516	105	0.2 (0.2, 0.3)	31,796	129	0.4 (0.3
Self-pay or Uninsured	219	2	0.9 (0.1, 3.3)	5,780	7	0.1 (0.0, 0.2)	0	0	N/A
Other (such as other government plans)	22	0	0.0 (0.0, 15.4)	2,927	20	0.7 (0.4, 1.1)	1,526	7	0.5 (0.2
Unknown	0	0	N/A	3	0	0.0 (0.0, 0.7)	576	2	0.3 (0.0

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

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

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

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

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

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

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

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

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

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

2b3.9. Results of Risk Stratification Analysis:

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

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

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

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

We examined the data to determine if there were meaningful differences in performance (harm rates) between measured entities (i.e., hospitals). We examined confidence intervals around the estimates and variation in performance rates between hospitals within **Beta Datasets 1, 2 and 3** to determine the stability of each estimate and if there were differences in performance (harm rates) between hospitals, respectively.

2b4.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

The performance rate across all hospitals in **Beta Datasets 1, 2, and 3** was 0.72% (95% CI: 0.68%, 0.76%). The performance ranged from 0.0% to 1.46% across all hospitals in **Beta Datasets 1, 2, and 3.**

The performance rate for all hospitals in **Beta Dataset 1** was 0.50 % (95% CI: 0.36%, 0.69%).

The performance rate for all hospitals in **Beta Dataset 2** was 0.72% (95% CI: 0.67%, 0.78%).

The performance rate for all hospitals in **Beta Dataset 3** was 0.73% (95% CI:0.67%, 0.81%).

2b4.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e.*, *what do the results mean in terms of statistical and meaningful differences?*)

Results from **Beta Datasets 1, 2, and 3** showed lower performance rates than rate of harm found in the literature, of 1.8% for all patients with a pressure injury, not only those acquired after 24 hours of hospitalization (Bauer 2016). This rate reflects data from chart abstraction, which is more burdensome for hospitals to collect. Because this measure includes only new pressure injuries, we anticipated the rates would be lower than those reported in the literature. However, the development of new pressure injuries, although rare, remains an important harm to assess. Additionally, there was variation in the rate of harm across the 24 hospitals in all Beta Testing datasets, demonstrating a quality signal and suggesting room for improvement in rates of hospital-acquired pressure injury among admitted patients.

References:

Bauer K, Rock K, Nazzal M, Jones O, Qu W. Pressure Ulcers in the United States' Inpatient Population From 2008 to 2012: Results of a Retrospective Nationwide Study. Ostomy Wound Manage. 2016;62(11):30-38.

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

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

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

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

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

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

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b6.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and 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 quantitatively assessed data element feasibility using the rate of missing for each required EHR data element for measure calculation.

For the EHR data elements used in this measure, we anticipate that there may be some missing data. However, we included only those variables that we expect to be consistently obtained in the target population, available in structured fields, and captured as part of the standard care workflow.

2b6.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

Table 1. Frequency of Missing Data by Data Element Required for the Denominator (Beta Datasets 1, 2, and 3)

Data Element	Beta Dataset 1			Beta Dataset 2			Beta Dataset 3		
	Missing Count (#)	Encounters (#)	Missing Percent (%)	Missing Count (#)	Encounters (#)	Missing Percent (%)	Missing Count (#)	Encounters (#)	Missing Percent (%)
Admission characteristic: Admission date and time	0	7,573	0.00%	0	100,238	0.00%	0	56,330	0.00%
Patient characteristic: Date of birth	0	7,573	0.00%	0	100,238	0.00%	0	56,330	0.00%

Table 8. Frequency of Missing Data by Data Element Required for Measure (Beta Datasets 1, 2, and 3)

Data Element	Beta Dataset 1			Beta Dataset 2			Beta Dataset 3		
	Missing Count (#)	Pressure Injuries (#)	Missing Percent (%)	Missing Count (#)	Pressure Injuries (#)	Missing Percent (%)	Missing Count (#)	Pressure Injuries (#)	Missing Percent (%)
Diagnosis: Pressure injury with date/time	N/A	N/A	N/A	0	3,522	0%	26	2,183	1.2%
Diagnosis: Pressure injury stage	N/A	N/A	N/A	0	3,522	0%	34	2,183	1.6%

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

Beta Dataset 1 did not have a field for pressure injuries in their EHR, only one for "wounds"; this meant we only used pressure injury data elements that had a stage attached, to ensure all data used were pressure injuries and not extraneous wounds. This health system is remedying their EHR currently.
Beta Datasets 2 and 3 had extremely low rates of missing data required for the measure calculation. We looked at the pressure injury-level instead of admission-level to get a clearer picture of the data reliability. The findings indicate that for health systems that are able to identify pressure injuries in discrete fields, all data elements required to calculate the measure are readily available and infrequently missing.

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

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

If other:

3b. Electronic Sources

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

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

ALL data elements are in defined fields in electronic 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. For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).

N/A; this is an eCQM that uses all data elements from defined fields in the electronic health record (EHR).

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

Attachment: Del18c3HOP5HarmsPressureInjuryFeasibilityScorecard12172018_v02-636824582773862567-636893854315492067.xlsx

3c. Data Collection Strategy

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

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

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

This measure is not instrument-based. As this measure is an eCQM and has not been implemented, difficulties with this measure have not been experienced. As noted above, feasibility assessment across 24 hospitals with three different EHR vendors showed that most data elements used to calculate the measure were reliably available in a structured format within the EHR, captured as part of the course of care, and coded using nationally accepted terminology.

3c.2. Describe any fees, licensing, or other requirements to use any aspect of the measure as specified *(e.g., value/code set, risk model, programming code, algorithm)*.

There are no fees associated with the use of this eCQM. Value sets are housed in the Value Set Authority Center (VSAC), which is provided by the National Library of Medicine (NLM), in coordination with the Office of the National Coordinator for Health Information Technology and the Centers for Medicare & Medicaid Services.

Viewing or downloading value sets requires a free Unified Medical Language System® (UMLS) Metathesaurus License, due to usage restrictions on some of the codes included in the value sets.

Individuals interested in accessing value set content can request a UMLS license at (https://uts.nlm.nih.gov/license.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.

Specific Plan for Use	Current Use (for current use provide URL)
Not in use	

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

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

N/A; this eCQM is under initial endorsement review and is not currently used in any accountability program.

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

N/A; this eCQM is under initial endorsement review and is not currently used in any accountability program. In December 2018, this eCQM was presented to the Measure Applications Partnership (MAP), who noted conditional support for rulemaking. Thus, CMS is considering implementation in an accountability program pending feedback received during NQF endorsement and rulemaking.

4a1.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

Following MAP's recommendations and support, we envision that this measure will be considered for accountability programs via future rulemaking.

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

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

N/A; this measure is being submitted as de novo as has not yet been implemented. Implementation is planned pending finalization of the NQF and CMS rulemaking processes.

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

N/A; this measure is being submitted as de novo as has not yet been implemented. Implementation is planned pending finalization of the NQF and CMS rulemaking processes.

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

Describe how feedback was obtained.

N/A; this measure is being submitted as de novo as has not yet been implemented. Implementation is planned pending finalization of the NQF and CMS rulemaking processes.

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

N/A; this measure is being submitted as de novo as has not yet been implemented. Implementation is planned pending finalization of the NQF and CMS rulemaking processes.

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

While this measure does not have usability information from measured entities, as it is being developed de novo and has not been implemented yet, our team sought input from multiple stakeholder groups throughout the measure development process. We believe in a transparent measure development process, and highly value the feedback received on the measure. During development, a technical expert panel composed of a variety of stakeholders was engaged at various stages of development to obtain balanced, expert input. We also solicited and received feedback on the measure through an MMS Blueprint 44-day Public Input Period during development.

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

As noted above, input received from TEP members was instrumental to the development and specification of this measure. Feedback received during public comment was also explored during the measure testing process.

Improvement

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

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

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

This is a new eCQM and there is no time trend information available regarding facility performance improvement. This eCQM is not currently used in any quality improvement program, but a primary goal of the eCQM is to provide hospitals with performance information necessary to implement focused quality improvement efforts.

4b2. Unintended Consequences

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

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

We did not identify any unintended consequences during eCQM development or testing. However, CMS is committed to monitoring this eCQM's use and assessing potential unintended consequences over time, such as the inappropriate shifting of care, and other negative unintended consequences for patients.

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

No unexpected benefits were noted during eCQM development 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)

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

Competing: Hospital-acquired pressure injuries are currently measured and publicly reported in the Hospital-Acquired Condition Reduction Program (HACRP) as a component of the Patient Safety Indicator (PSI) 90 measure, which relies on ICD codes as a data source.

Related: Additionally, the following NQF endorsed measures are related but measure different patient populations: Percent of High Risk Residents with Pressure Ulcers (Long Stay) (NQF #0679) and Percent of Residents or Patients with Pressure Ulcers that are New or Worsened (Short Stay) (NQF #0678).

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

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

Are the measure specifications harmonized to the extent possible?

Yes

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

While there are several measures that target the reduction of hospital-acquired pressure injuries in use in various patient populations, there are no eCQMs intended for use to compare quality across acute care hospitals. The measures NQF# 0679 and #0678 target a different patient population and use chart review data from the following

sources: Minimum Data Set (MDS); Long Term Care Hospitals Continuity Assessment Record and Evaluation (LTCH-CARE) Data set; and the Inpatient Rehabilitation Facility Patient Assessment Instrument (IRF-PAI) Data set. Additionally, NQF# 0678 measure includes worsening pressure injuries and NQF# 0679's population consists of only high-risk patients defined as those who are impaired in bed mobility, comatose, or suffering malnutrition. The new Hospital Harm -Pressure Injury eCQM identifies pressure injuries using direct extraction of structured data from the EHR and will provide hospitals with reliable and timely measurement of their pressure injury rates. As these measures do not apply to the same measured entities, it should not impact 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.)

Hospital-acquired pressure injuries are currently measured and publicly reported in the Hospital-Acquired Condition Reduction Program (HACRP) as a component of the Patient Safety Indicator (PSI) 90 measure (PSI-03). PSI-03 does not include stage 2 pressure injuries in the outcome, has additional exclusions to the cohort, and uses ICD codes via claims as a data source. Hospital Harm - Pressure Injury Measure is an eCQM (EHR data-only), which stakeholders and TEP have noted as a more desirable data source with more face validity for measuring pressure injuries.

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:** Pressure_Injury_Algorithm.docx

Contact Information

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

Co.2 Point of Contact: Joseph, Clift, Joseph.Clift@cms.hhs.gov, 410-786-4165-

Co.3 Measure Developer if different from Measure Steward: IMPAQ International LLC

Co.4 Point of Contact: Benjamin, Shirley, bshirley@impaqint.com, 202-774-1964-

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.

Technical Expert Panel Members

David Baker, MD, MPH, The Joint Commission

Cynthia Barnard, PhD, MBA, MSJS, Northwestern Memorial Healthcare

Lisa Freeman, BA, Connecticut Center for Patient Safety

Patrick Guffey, MD, University of Colorado Department of Anesthesiology

David Hopkins, MS, PhD, Stanford University School of Medicine

Kevin Kavanagh, MD, MS, Health Watch USA

Joseph Kunisch, PhD, RN-BC, CPHQ, Memorial Hermann Hospital System

Timothy Lowe, PhD, Premier Inc.

Christine Norton, MA, Patient/Consumer/Caregiver

Amita Rastogi, MD, MHA, CHE, MS, Remedy Partners

Karen Zimmer, MD, MPH, Jefferson School of Population Health

Julia Hallisy, The Empowered Patient Coalition (served from March 2017 to September 2017)

Jennifer Meddings, MD, MSc, University of Michigan Health System (served from March 2017 to October 2018)

Eric Thomas, MD, MPH, McGovern Medical School at University of Texas Health (served from March 2017 to October 2018)

Technical Advisory Group Members

Andy Anderson, MD, MBA, RWJ Barnabas Health and Rutgers University

Matt Austin, MS, PhD, John Hopkins Medicine

Ann Borzecki, MD, Department of Veteran´s Affairs

John Bott, The Leapfrog Group

Kyle Bruce, DPM, Riverbend Medical Group

David C. Chang, PhD, MPH, MBA, Massachusetts General Hospital, Harvard Medical School

Hazel R. Crews, MHA, MHS, CPHQ, Indiana University Health

Melissa Danforth, The Leapfrog Group

Richard Dutton, MD, Baylor University

Marybeth Foglia, RN, PhD, MA, National Center for Ethics in Healthcare

Jeff Giullian, MD, MBA, DaVita Kidney Care

Maryellen Guinan, America´s Essential Hospitals

Kate Kovich, Advocate Health Care

David Levine, MD, FACEP, Vizient Center for Advanced Analytics and Informatics

Karen Lynch, E, RN MGH, LCS, Massachusetts General Hospital

Milisa Manojlovich, MD, University of Michigan

Barbara Pelletreau, Dignity Health

Marc Philip Pimentel, T.M.D., Brighham and Women´s Hospital

Christine Sammer, DrPH, RN, CPPS, FACHE, Adventist Health System

Brett Stauffer MD MHS FHM, Baylor Scott and White Health

Brooks Udelsman, MD/MHS, Massachusetts General Hospital

Boback Ziaeian, UCLA

Similar to our TEP, these experts responded to the posted Call for TEP members. The Technical Advisory Group was utilized similar to a TEP, providing feedback on clinical acceptability of measure specifications and feasibility of the measure.

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? As a de novo measure submission, we anticipate annual updates and potentially triennial endorsement.

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

Ad.6 Copyright statement: Limited proprietary coding is contained in the measure specifications for convenience. Users of the proprietary code sets should obtain all necessary licenses from the owners of these code sets CPT(R) contained in the Measure specifications is copyright 2004-2016 American Medical Association. LOINC(R) copyright 2004-2016 Regenstrief Institute, Inc. This material contains SNOMED Clinical Terms(R) (SNOMED CT[R]) copyright 2004-2016 International Health Terminology Standards Development Organisation. ICD-10 copyright 2016 World Health Organization. All Rights Reserved.

Ad.7 Disclaimers: This measure and specifications are subject to further revisions. This performance measure is not a clinical guideline and does not establish a standard of medical care, and has not been tested for all potential applications. THE MEASURES AND SPECIFICATIONS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND. Due to technical limitations, registered trademarks are indicated by (R) or [R] and unregistered trademarks are indicated by (TM) or [TM].

Ad.8 Additional Information/Comments: This measure was originally developed, specified, and tested by YALE CORE and Mathematica Policy Research on behalf of the Centers for Medicare and Medicaid Services (CMS). IMPAQ International, LLC assumed developer responsibility for this measure in March 2019.