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 #: 3533e

Corresponding Measures:

De.2. Measure Title: Hospital Harm – Severe Hyperglycemia

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

De.3. Brief Description of Measure: This ratio electronic clinical quality measure (eCQM) assesses the number of hospital days with a severe hyperglycemic event (a blood glucose result >300 mg/dL, or a day in which a blood glucose value was not documented and it was preceded by two consecutive days where at least one glucose value is >=200 mg/dL) per the total qualifying hospital days among inpatient encounters for patients 18 years and older who have either:

1. A diagnosis of diabetes mellitus,

2. Received at least one administration of insulin or an anti-diabetic medication during the hospital admission, or

3. Had an elevated blood glucose level (>200 mg/dL) during their hospital admission.

1b.1. Developer Rationale: This eCQM relates to glycemic management in the hospital inpatient setting. Rates of inpatient severe hyperglycemic events – an extremely elevated blood glucose level – can be considered an indicator of quality of care provided by a hospital. Severe hyperglycemia is associated with a range of harms, including increased in-hospital mortality, infection rates, and hospital length of stay.1,2,3,4,5,6,7,8 The rate of severe hyperglycemic events varies across hospitals, which suggests that there are opportunities for improvement in glycemic management.9,10,11 The implementation of this eCQM will aim to achieve several improvements in quality. For instance, this eCQM will encourage providers to develop interventions aimed at better glycemic control and prevent severe hyperglycemia for hospital inpatients. In addition to avoiding direct patient harm from the severe hyperglycemic event, lower rates of severe hyperglycemia among hospitalized individuals would be expected to result in lower rates of mortality, infection, and hospital length of stays. Adopting this eCQM has the potential to improve quality of care for individuals at risk of hyperglycemia and, therefore, advance the quality of care in patient safety, which is a priority area identified by the National Quality Strategy. This will fill a gap in measurement and provide incentives for hospital quality improvement, as there is no hyperglycemia measure in any CMS program. With a systematic EHR-based patient safety measure in

place, hospitals can more reliably assess harm reduction efforts and modify their improvement efforts in near real-time. In addition, we can expect to make greater achievements in reducing harms and enhancing hospital performance on patient safety outcomes.

References:

1. Pasquel FJ, Spiegelman R, McCauley M, et al. Hyperglycemia During Total Parenteral Nutrition: An Important Marker of Poor Outcome and Mortality in Hospitalized Patients. Diabetes Care. 2010;33(4):739-741

2. Rady MY, Johnson DJ, Patel BM, Larson JS, Helmers RA. Influence of Individual Characteristics on Outcome of Glycemic Control in Intensive Care Unit Patients With or Without Diabetes Mellitus. Mayo Clin Proc. 2005;80(12):1558-1567.

3. Umpierrez GE, Isaacs SD, Bazargan N, You X, Thaler LM, Kitabchi AE. Hyperglycemia: An Independent Marker of In-Hospital Mortality in Patients with Undiagnosed Diabetes. J Clin Endocrinol Metab. 2002;87(3):978-982.

4. Falciglia M, Freyberg RW, Almenoff PL, D´Alessio DA, Render ML. Hyperglycemia-Related Mortality in Critically Ill Patients Varies with Admission Diagnosis. Crit Care Med. 2009;37(12):3001-3009.

5. Lee LJ, Emons MF, Martin SA, et al. Association of Blood Glucose Levels with In-Hospital Mortality and 30-Day Readmission in Patients Undergoing Invasive Cardiovascular Surgery. Curr Med Res Opin. 2012;28(10):1657-1665.

6. King JT, Jr., Goulet JL, Perkal MF, Rosenthal RA. Glycemic Control and Infections in Patients with Diabetes Undergoing Noncardiac Surgery. Ann Surg. 2011;253(1):158-165.

7. Jackson RS, Amdur RL, White JC, Macsata RA. Hyperglycemia is Associated with Increased Risk of Morbidity and Mortality after Colectomy for Cancer. J Am Coll Surg. 2012;214(1):68-80.

8. Umpierrez GE, Hellman R, Korytkowski MT, et al. Management of Hyperglycemia in Hospitalized Patients in Non-Critical Care Setting: An Endocrine Society Clinical Practice Guideline. J Clin Endocrinol Metab. 2012;97(1):16-38.

9. Swanson CM, Potter DJ, Kongable GL, Cook CB. Update on Inpatient Glycemic Control in Hospitals in the United States. Endocr Pract. 2011;17(6):853-861.

10. Cook CB, Kongable GL, Potter DJ, Abad VJ, Leija DE, Anderson M. Inpatient Glucose Control: A Glycemic Survey of 126 U.S. Hospitals. J Hosp Med. 2009;4(9):E7-E14.

11. Matheny ME, Shubina M, Kimmel ZM, Pendergrass ML, Turchin A. Treatment Intensification and Blood Glucose Control among Hospitalized Diabetic Patients. J Gen Intern Med. 2008;23(2):184-189.

S.4. Numerator Statement: The total number of hyperglycemic days across all encounters divided by the total number of eligible days across all encounters. Hospital days are measured in 24-hour periods, starting from the time of arrival at the hospital (including Emergency Department). Days with a hyperglycemic event are defined as:

- A day with at least one blood glucose value >300 mg/dL; or

- A day in which a blood glucose value was not documented and it was preceded by two consecutive days where at least one glucose value is >=200 mg/dL.

We do not count >300 mg/DL events the first 24-hour period after admission to the hospital (including the Emergency Department) or the last time period before discharge, if it was less than 24 hours.

S.6. Denominator Statement: The initial population is all patients 18 years and older at the start of the measurement period with a discharged inpatient hospital admission during the measurement period, as well as either:

1. A diagnosis of diabetes that starts before or during the encounter; or

2. Administration of at least one dose of insulin or any anti-diabetic medication during the encounter; or

3. Presence of at least one blood glucose value >200 mg/dL at any time during the encounter.

The eCQM includes inpatient encounters which began in the Emergency Department or in observation status.

The denominator is the total number of eligible days across all encounters which match the initial population criteria. We do not count the the first 24-hour period after admission to the hospital (including the Emergency Department) or the last time period before the discharge, if it was less than 24 hours. By excluding the first 24 hours of admission, we allow for correction of severe hyperglycemia that was present on admission. By excluding the last time period before discharge if it was less than 24 hours, we account for the fact that hospitals may not always be able to check glucose during the last time period, especially if it is only a few hours long. Eligible encounters that exceed 10 days are truncated to equal 10 days.

S.8. Denominator Exclusions: N/A; 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? This eCQM is intended to be used simultaneously with the eCQM assessing severe hypoglycemia events (NQF#3503e) during acute short-term care admissions for patients age 18 years and older. The Hospital Harm - Severe Hypoglycemia eCQM functions as a balancing measure for this eCQM by minimizing the potential for unintended consequences such as fluctuating glycemic values and hypoglycemia.

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 logic model presented by the developer for this outcome measure links evidence based standardized protocols and insulin management protocols with improved glycemic control (lower rates of severe hyperglycemic events) and safety outcomes (hospital mortality, infection rates, lower hospital length of stay).
- Per developer, the goal of this eCQM is to improve patient safety and prevent severe hyperglycemia in patients during their hospitalization. The developer submitted studies indicating severe hyperglycemia can be reduced through proper glycemic management.
- The developer also noted literature that the rate of severe hyperglycemia varies across hospitals, suggesting opportunities for improvement in care. Per developer, hyperglycemic rates have been reported from 32.2% to 46.0% of ICU patient-days, and 31.7% to 54.2% of non-ICU patient-days (>180 mg/dL).

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

Does the measure assess performance on a health outcome (Box 1) -> (yes) -> Is there a relationship between the measure and at least one healthcare action is demonstrated by empirical data (Box 2) -> (yes) -> PASS

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 seven hospitals in four regions (West, Midwest, Southeast, South). Hospitals varied in size (100-799 beds) and EHR systems (Cerner, Meditech, Epic). Six of the seven hospitals were teaching hospitals, and two were located in rural areas while five hospitals were located in urban areas. Data were collected 1/1/2018-3/31/2018 for hospital 1, 2, 3, and 4. Data were collected 1/1/2018-10/31/2018 for hospital 5, 6, and 7.
- The range in performance across tested hospitals was from 8.2% to 19.5%. The overall performance rate for six of the hospitals was 13.6%. The 95% confidence interval was 13.1%, 14.1.%. SD 0.2%. Developer also provided individual performance rates for the six hospital. No performance rate for hospital 7 was calculated due to inability to map POC glucose lab data at time of testing.
- The rate of severe hyperglycemic events varies across the six hospitals, which suggests that there are opportunities for improvement in glycemic management.

Disparities

- The developer did not cite literature. However, the measure performance was stratified by age/gender/race/ethnicity (collected 1/1/2018-3/31/2018 for hospital 1, 2, 3, and 4 and was collected 1/1/2018-10/31/2018 for hospital 5 and 6). The developer noted some limits on usability of results of race/ethnicity due to some facilities identified it as an "unknown".
- Statification of measure performance by age (Age//Denominator//Numerator//**Measure Rate** (95% Confidence Interval))
    - 18-64//7,471//1,128//**15.1%** (14.1%, 16.1%)
    - 65+//12,265//1,551//**12.6%** (12.1%, 13.3%)
- Statification of measure performance by gender (Gender//Denominator//Numerator//**Measure Rate** (95% Confidence Interval))
    - Male//10,037 //1,265//**12.6%** (12.0%, 13.3%)
    - Female//9,699//1,414//**14.6%** (13.9%, 15.3%)

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *Are you aware of evidence that disparities exist in this area of healthcare?*
- *Is there any need for risk-stratification of this measure?*

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

RATIONALE: The rate of severe hyperglycemic events varies across hospitals, which suggests that there are some opportunities for improvement in glycemic management.

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

1a. Evidence

Comments:

**no

**These data and evidence are more mature now; the emeasure does not assess isolated elevated values, but rather very high measures that are repeat and outside the early care pahses.

**no

**outcome measure and is a good proxy for missed diagnosis

**There is a clear link between the measure and an important health outcome.

**Solid evidence base

1b. Performance Gap

Comments:

**yes

**Measure stewards show ebvidence of gaps that vary across sites and populations.

**A substantial gap was reported in ICU and non-ICU patients

**yes, there is a gap

**Based on the data provided, there does appear to be a persistent performance gap. It is not clear to me exactly what the overall size of the population is based on such a small sample (6 hospitals with complete data – however the denominators are substantial) but the range of performance from 8.2-19.5% suggests an opportunity for improvement.

**Variability among facilities.

Disparities:

**no

**See above - while variability exists among subpopulations, the variability and disparity opportunity is similar to or smaller than other care processes.

**I did not see disparities by racial groups

**Disparities data is shown but does not show a significant gap

**It is challenging to assess disparities with only 6 hospitals (and thus 6 geographies) represented.

**Not known

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.

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

Evaluators: NQF Scientific Methods Panel Subgroup

Methods Panel Review (Combined)

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

Scientific Methods Panel Votes: Measure passes

- Reliability: H-6, M-0, L-0, I-0 → Measure passes with HIGH rating
- Validity: H-4, M-1, L-0, I-1 → Measure passes with HIGH rating

This measure was reviewed by the Scientific Methods Panel. A summary of the measure is provided below:

Reliability

- To assess reliability of the measure score, the developers used Adams' beta-binomial method to calculate the signal-to-noise ratio.
    - There were 5,501 eligible encounters (and 19,736 eligible days) across Hospitals 1-6. The signal-to-noise ratio yielded a median reliability score of 0.967 (range: 0.955-0.983). This is nearly perfect (the highest reliability score is 1.0).

Validity

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

NQF eMeasure Evaluation Summary:

Sub-criterion 2a1 - Specifications

- Submitted measure specification follows eCQM industry specs as indicated Sub-criterion 2a1
- Submitted measure specification is fully represented and is not hindered by any limitations in the eCQM industry specs

Sub-criterion 2a2 - Reliability

- Submission includes simulated data set results demonstrating unit testing covering 100% of the measure logic.

Sub-criterion 2b1 - Validity testing demonstrates that the measure data elements are correct

- no issues

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**

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

2a1. Reliability – Specifications

Comments:

**none

**The tragets are well defined and clear, accessible in e-record. I have no concerns.

**No concerns

**there is a complicated logic applied to some data (specifically if blood work not followed up upon) that may cause a problem

**The data elements (numerator, denominator and exclusions) are clearly defined. The reliability of the data elements which are drawn from the EMR appear reliable.

**Good reliability

2a2. Reliability – Testing

Comments:
**no
**No
**No
**only whether the query that is used can be easily reproduced
**No - based on reliability testing, this appears to be a very reliable measure (free from measurement error) with a signal to noise ratio >0.9
**no concerns

2b1. Validity –Testing

Comments:

**no

**None - see above - no malalignment with mesaure instructions by NQF.

**NO

**no

**No- there appears to strong evidence for the validity based on the robust analysis provided of the manual chart abstraction and the eCQM.

**no concerns

2b4-7. Threats to Validity
2b4. Meaningful Differences

Comments:

**none

**No - well considered and likely a very small value based on commonality of the test result/patient factors sought.

**None

**no

**None

**none evident

2b2-3. Other Threats to Validity
2b2. Exclusions
2b3. Risk Adjustment

Comments:

**not risk adjusted

**No adjusting or significant exclusions aside from Dx of diabetes or therapy for elevated blood sugar.

**I saw no risk adjustment and it may not be needed.

**I think one aspect is whether the hyperglycemia is a result of another condition or due to missed medication dosing

**The lack of risk adjustment does raise concern that the variation in performance could be accounted for by case mix. At the core of the developers argument is that "harms such as severe hyperglycemia are avoidable, regardless of patient risk." This is a challenging argument. The developers go on to describe their rational – which acknowledges the different challenges inherent in caring for different patient populations – but rests on the idea that with enough intense focus, all patients can achieve euglycemia while in the hospital. Their clinical reasoning is well laid out (p56-57), however it will be important to further investigate once there is a larger sample available.

**No risk adjustment, virtually all patients can achieve the goals

Combined Methods Panel Scientific Acceptability Evaluation

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 3533e

Measure Title: Hospital Harm – Severe Hyperglycemia

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.**

   Panel Member #1: None.

   Panel Member #2: None

   Panel Member #4: No Concerns

   Panel Member #5: No concerns

   Panel Member #6: None

RELIABILITY: TESTING

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

3. **Reliability testing level** ☒ **Measure score** ☐ **Data element** ☐ **Neither**
4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure** ☒ **Yes** ☐ **No**

5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical VALIDITY testing** of patient-level data conducted?

   ☐ **Yes** ☐ **No**

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

   Submission document: Testing attachment, section 2a2.2

   Panel Member #1: **Used Adam's beta binomial to assess signal-to-noise; appropriate method.**

   Panel Member #2: Signal-to-noise ratio as measure of reliability.

   Panel Member #3: Adam's binoial method, signal-to-noise

   Panel Member #4: No concerns, the Adam's beta-binomial method was appropriate.

   Panel Member #5: The developers estimated signal-to-noise reliability using the beta-binomial model. This model may not be a literally correct description of the data (e.g. because it ignores dependence between multiple hospital days from the same patient) but is acceptable to me for practical reasons due to lack of well-defined alternative methods. A limitation is that beta-binomial parameters were estimated from only ~6 hospitals and may not be generalizable to the entire hospital population of interest.

   Panel Member #6: The developer used a signal to noise analysis which is appropriate for this application.

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

   Submission document: Testing attachment, section 2a2.3

   Panel Member #1: **High values; median value of 0.967 (range: 0.955-0.983)**

   Panel Member #2: Median reliability 0.97, which is considered high reliability.

   Panel Member #3: Almost perfect agreement 0.96

   Panel Member #4: No concerns, demonstrated significant score reliability

   Panel Member #5: The median estimated reliability across all hospital sample sizes was 0.967.

   Panel Member #6: The results indicate that the measure is relatively free from measurement error, with a median reliability estimate of .97 across six hospitals. Note that the Landis & Koch rubric is for inter-rater reliability, not signal to noise. Regardless, the reliability of this measure is high.

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

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

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

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

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

☐ **No**

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

10. **OVERALL RATING OF RELIABILITY** (taking into account precision of specifications and all testing results):

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

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

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

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

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

    Panel Member #1: Score-level testing; appropriate method; excellent results.

    Panel Member #2: See reposne in #7 above.

    Panel Member #3: No concerns

    Panel Member #4: No concerns

    Panel Member #5: The reported data suggest wide signal variation and large hospital-specific sample sizes.

    Panel Member #6: The range of signal to noise estimates indicates very high reliability across the six hospitals (range of .95 - .98).

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

    Submission document: Testing attachment, section 2b2.

    Panel Member #1: Not applicable.

    Panel Member #2: N/A

    Panel Member #3: N/A

    Panel Member #4: No concerns, exclusions using the initial admit time peiod removes patients with high glucoses from the initial patient population appropriately.

    Panel Member #5: None

    Panel Member #6: N/A

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

    Submission document: Testing attachment, section 2b4.

    Panel Member #1: My only comment is in addition to the overall statistics, it would be helpful to see results for all six hospitals with their associated confidence intervals.

Panel Member #2: None

Panel Member #3: Showed variation in rates across hospitals; however, did not offer an interpretation of the degree of impact.

Panel Member #4: No concerns

Panel Member #5: None

Panel Member #6: None. There was good variation in the outcome among this small sample of hospitals.

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.
Panel Member #1: **Does not apply.**
Panel Member #2: N/A
Panel Member #3: None
Panel Member #6: N/A

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

Submission document: Testing attachment, section 2b6.

Panel Member #1: **None. Only looked at missing data for 3 hospitals, but the three hospitals were quite different from each other (rural/urban, teaching vs. non-teaching).**

Panel Member #2: None

Panel Member #4: No concerns

Panel Member #5: None

Panel Member #6: 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

Panel Member #1: **(Argue that all patient can be managed to stay below 300).**

Panel Member #5: (Checking the 'yes' box because the developers provided a rationale, not because I necessarily agree with the lack of risk adjustment.)

16c. **Social risk adjustment:**

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

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

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

16d.**Risk adjustment summary:**

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

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

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

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

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

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

Panel Member #2: Agree with the developer rationale for not doing any risk-adjustment.

Panel Member #4: Submitors provide clinically based rationale for no risk adjustment but I would encourage them to investigate risk adjustment when more data is available as there is significant variation in performance in the patient characteristic tables that may impact overall performance.

Panel Member #5: Due to no risk adjustment, observed performance differences across hospitals could be at least partly explained by case mix.

For cost/resource use measures ONLY:

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

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

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

VALIDITY: TESTING

19. **Validity testing level: ☐ Measure score ☒ Data element ☒ Both**

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

☒ Face validity

☒ Empirical validity testing of the measure score

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

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

Submission document: Testing attachment, section 2b2.2

Panel Member #1: Face validity (score-level): queried 11 experts, asking if measure would differentiate quality (+)

Emprical testing (score-level): it was unclear to me how their score-level testing was different from their data element testing (-)

Empirical testing (data element): electronically extracted and compared EHR data to chart review (+)

Panel Member #2: Face validity through TEP;

Data element validity through assessing the accuracy of EHR-extracted data with manual chart abstraction;

Measure score validty through PPV, sensitivity, specificity and NPV.

Panel Member #3: Compared EHR data elements with abstracted elements

PPV, chart rate score with measure score

TEP

Panel Member #4: No concerns

Panel Member #5: Empirical analyses focused on assessing accuracy of the electronic EHR extraction algorithm for correctly classifying a patient's status (hyperglycemic or not) on a day to day basis. The developer's describe this as score-level validity but arguably it is targeting element-level validity. The developers argue that comparison to an external measure of quality was unnecessary because the measure is not risk adjusted and because the measure is simply summing up the number of true harm events. The argument is somewhat circular as the phrase "harm" implicitly assumes that hyperglycemic events are attributable to the measured entity e.g. as opposed to random sampling variation, case mix, etc. and that they are a relevant and valid outcome to measure. On the other hand, I don't feel that score-level empirical testing can directly address these issues so I am not advocating to require them.

Panel Member #6: Kappa scores were provided but only for the numerator condition of 'severe hyperglycemia'. However, the denominator conditions (diabetes diagnosis, admin of insulin/anti-diabetic medication, low glucose reading as well as numerator elements such as admission date were not evaluated. Face validity methodology was not described, therefore cannot determine if it was systematic or transparent.

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

Submission document: Testing attachment, section 2b2.3

Panel Member #1: **Face validity (score-level): 10/11 experts agreed that measure would differentiate quality**

Empirical testing (data element): 100% sensitivity

Panel Member #2: Measure validity was established across all the above three components of validity (face validity, data element validity and score-level validity).

Panel Member #3: High sensitivity, PPV, TEP member confirmation

Panel Member #4: No concerns, submitors demonstrated strong PPV and adequate face validity

Panel Member #5: SMO: The results indicate that the EHR algorithm had high accuracy for classifying hyperglycemic days.

Panel Member #6: Severe hyperglyemica definition had strong agreement, but other critical data elements of the measure were not tested.

23. **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)

24. **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)

25. **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.)

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

    Panel Member #1: **Tested both score-level and data elements; good results on both.**

    Panel Member #2: See note in #22 above.

    Panel Member #3: Multi-method approach

    Panel Member #4: No concerns, submitors demonstrated significant statistical validity along with face validity. Sensitivity and specificity were close to perfect.

    Panel Member #6: Either more information on the face validity results are needed (how many TEP members, who were the TEP members, how were the TEP members assessed) or kappa results for all data elements are needed to assess validity.

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 Specifications and Elements

- The measure is generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score)
- ALL data elements are in defined fields in electronic health records (EHRs)
- This measure is an eCQM.

Data Collection Strategy

- This measure is a new measure and not in use currently, so the developer did not provide any have information on difficulties with data collection.
- There are no fees associated with the use of this eCQM.

NQF eMeasure Feasibility Evaluation Summary:

Sub-criterion - 3a the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

- no issues

Sub-criterion 3b - 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.

- Date element Laboratory Test, Performed: Gluose Lab Test was assessed as having feasibility issues in the availability/standards domain(s) indicating that the data element may not be available electronically or have a credible near term path to electronic collection.
- The developer provided additionl clarification to NQF in December 2019:
    - "We would like to clarify that testing did demonstrate that the data element, Laboratory Test, Performed: Glucose Lab Test, was documented and retrievable from multiple sites and multiple EHRs as indicated on the feasibility scorecard. The challenge, as in many eCQMs, was that the vendor and/or local codes used at one hospital with one vendor were not already mapped to the LOINC codes in the defined value set. Going forward, similar to workflow processes for other eCQMs in CMS programs, the Glucose Lab Tests from local and vendor codes would be appropriately mapped to correct LOINC codes in the measure value set."
    - "Please note that although terminology standards for all data elements are currently available, the LOINC standards were not consistently coded to that standard terminology for one data element, Laboratory Test Performed: Glucose Lab Test, at one site. Mapping the local vendor codes to the appropriate LOINC codes is credible near-term path to electronic collection for that particular data element at that particular site."

Sub-criterion 3c - Feasibility

- All value sets used in measure submission are accessible via the VSAC
- Submission includes simulated data set results demonstrating unit testing covering 100% of the measure logic.

Questions for the Committee:

- Are the required data elements routinely generated and used during care delivery?
- Are the required data elements available in electronic form, e.g., EHR or other electronic sources?
- Is the data collection strategy ready to be put into operational use?

- Does the eCQM Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites?

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

RATIONALE:

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

3. Feasibility

Comments:

**none

**No concerns - the data needed are easuily accessed and reliably so thru surveillence methods that are non-manual.

**No

**the query has to be tried on a hospital level, though the EMR vendors could create it

**Concerned about feasibility given that 1/7 hospitals could not retrieve the data. Comment from developers that "With incentive from CMS (by adding to a rule and requiring implementation), the mapping would be completed by EHRs vendors in advance and would thus enable full implementation at the organization in question" raises concern.

**no concerns

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

- The measure is not currently in use. This measure is being developed for the Hospital Inpatient Quality Reporting (HIQR) and the Promoting Interoperability (PI) for Eligible Hospitals and

Critical Access Hospitals programs pending NQF endorsement, Measure Application Partnership (MAP) pre-rulemaking evaluation, and the CMS rulemaking process.

4a.2. Feedback on the measure by those being measured or others. Three criteria demonstrate feedback: 1) those being measured have been given performance results or data, as well as assistance with interpreting the measure results and data; 2) those being measured and other users have been given an opportunity to provide feedback on the measure performance or implementation; 3) this feedback has been considered when changes are incorporated into the measure

Feedback on the measure by those being measured or others

- The measure is currently not implemented in a public reporting or accountability program. Currently feedback has been via their Technical expert panel, as well as the MMS Blueprint public comment period on this measure.

Additional Feedback:

- The measure is also submitted to the Measure Application Partnership (MAP) on the November 2019 Measures Under Consideration List (MUC 19-26) to the NQF MAP Hospital Workgroup, which convened in December 2019.

***Questions for the Committee**:*

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

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

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 This eCQM is new and not currently in use in any quality improvement programs.

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 There are no unexpected findings indentified by the developer.

Potential harms There are no harms identified by the developer.

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:

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

4a1. Use - Accountability and Transparency

Comments:

**none

**No clear use by others described, but much opportunity exists.

**No

**This would be a complicated measure per se to make public - ie how to explain significance, though it is significant

**Public reporting not currently taking place, although a credible plan for implementation (being developed for HIQR) is presented.

**measure not in use yet, feedback in development process

4b1. Usability – Improvement

Comments:

**none

**Limited harm data though none anticipated - that has face validity.

**It's not clear to me how the threshold of 300 was chosen in view of potential harm from hyperglycemia and risk of hypoglycemia if over-control results.

**no harm to it

**No concerns

**no concerns

Criterion 5: Related and Competing Measures

Related or competing measures

- There are currently no NQF endorsed hyperglycemia measures. NQF recently endorsed 3503e Hospital Harm-Severe hypoglycemia measure by the same steward/developer as 3533e.

Harmonization

- Per developer, this eCQM is intended to be used simultaneously with the eCQM assessing severe hypoglycemia events (NQF#3503e) during acute short-term care admissions for patients age 18

years and older. The Hospital Harm - Severe Hypoglycemia eCQM functions as a balancing measure for this eCQM by minimizing the potential for unintended consequences such as fluctuating glycemic values and hypoglycemia.

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

5. Related and Competing

Comments:

**none

**3503e is a maesure intended to be used together with this measure and is aligned.

**None that I am aware of.

**not that I am aware

**The only related measure that is worth discussing (other than the developers related hypoglycemia measure which seems fairly complimentary with no need for further alignment) is 2362e which was retired by CMS. It appears the key changes are that there has been a change in the value defined as hyperglycemic (from 200 to 300), the new measure now includes patients admitted for DKA/HHS, there is no risk adjustment, and the new measures accounts for LOS. Is there any data on why the previous measure was retired?

**hypoglycemia measure to avoid unintended harms

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 1/21/2020

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

Developer Submission

Brief Measure Information

NQF #: 3533e

Corresponding Measures:

De.2. Measure Title: Hospital Harm – Severe Hyperglycemia

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

De.3. Brief Description of Measure: This ratio electronic clinical quality measure (eCQM) assesses the number of hospital days with a severe hyperglycemic event (a blood glucose result >300 mg/dL, or a day in which a blood glucose value was not documented and it was preceded by two consecutive days where at least one glucose value is >=200 mg/dL) per the total qualifying hospital days among inpatient encounters for patients 18 years and older who have either:

1. A diagnosis of diabetes mellitus,

2. Received at least one administration of insulin or an anti-diabetic medication during the hospital admission, or

3. Had an elevated blood glucose level (>200 mg/dL) during their hospital admission.

1b.1. Developer Rationale: This eCQM relates to glycemic management in the hospital inpatient setting. Rates of inpatient severe hyperglycemic events – an extremely elevated blood glucose level – can be considered an indicator of quality of care provided by a hospital. Severe hyperglycemia is associated with a range of harms, including increased in-hospital mortality, infection rates, and hospital length of stay.1,2,3,4,5,6,7,8 The rate of severe hyperglycemic events varies across hospitals, which suggests that there are opportunities for improvement in glycemic management.9,10,11 The implementation of this eCQM will aim to achieve several improvements in quality. For instance, this eCQM will encourage providers to develop interventions aimed at better glycemic control and prevent severe hyperglycemia for hospital inpatients. In addition to avoiding direct patient harm from the severe hyperglycemic event, lower rates of severe hyperglycemia among hospitalized individuals would be expected to result in lower rates of mortality, infection, and hospital length of stays. Adopting this eCQM has the potential to improve quality of care for individuals at risk of hyperglycemia and, therefore, advance the quality of care in patient safety, which is a priority area identified by the National Quality Strategy. This will fill a gap in measurement and provide incentives for hospital quality improvement, as there is no hyperglycemia measure in any CMS program. With a systematic EHR-based patient safety measure in place, hospitals can more reliably assess harm reduction efforts and modify their improvement efforts in near real-time. In addition, we can expect to make greater achievements in reducing harms and enhancing hospital performance on patient safety outcomes.

References:

1. Pasquel FJ, Spiegelman R, McCauley M, et al. Hyperglycemia During Total Parenteral Nutrition: An Important Marker of Poor Outcome and Mortality in Hospitalized Patients. Diabetes Care. 2010;33(4):739-741

2. Rady MY, Johnson DJ, Patel BM, Larson JS, Helmers RA. Influence of Individual Characteristics on Outcome of Glycemic Control in Intensive Care Unit Patients With or Without Diabetes Mellitus. Mayo Clin Proc. 2005;80(12):1558-1567.

3. Umpierrez GE, Isaacs SD, Bazargan N, You X, Thaler LM, Kitabchi AE. Hyperglycemia: An Independent Marker of In-Hospital Mortality in Patients with Undiagnosed Diabetes. J Clin Endocrinol Metab. 2002;87(3):978-982.

4. Falciglia M, Freyberg RW, Almenoff PL, D´Alessio DA, Render ML. Hyperglycemia-Related Mortality in Critically Ill Patients Varies with Admission Diagnosis. Crit Care Med. 2009;37(12):3001-3009.

5. Lee LJ, Emons MF, Martin SA, et al. Association of Blood Glucose Levels with In-Hospital Mortality and 30-Day Readmission in Patients Undergoing Invasive Cardiovascular Surgery. Curr Med Res Opin. 2012;28(10):1657-1665.

6. King JT, Jr., Goulet JL, Perkal MF, Rosenthal RA. Glycemic Control and Infections in Patients with Diabetes Undergoing Noncardiac Surgery. Ann Surg. 2011;253(1):158-165.

7. Jackson RS, Amdur RL, White JC, Macsata RA. Hyperglycemia is Associated with Increased Risk of Morbidity and Mortality after Colectomy for Cancer. J Am Coll Surg. 2012;214(1):68-80.

8. Umpierrez GE, Hellman R, Korytkowski MT, et al. Management of Hyperglycemia in Hospitalized Patients in Non-Critical Care Setting: An Endocrine Society Clinical Practice Guideline. J Clin Endocrinol Metab. 2012;97(1):16-38.

9. Swanson CM, Potter DJ, Kongable GL, Cook CB. Update on Inpatient Glycemic Control in Hospitals in the United States. Endocr Pract. 2011;17(6):853-861.

10. Cook CB, Kongable GL, Potter DJ, Abad VJ, Leija DE, Anderson M. Inpatient Glucose Control: A Glycemic Survey of 126 U.S. Hospitals. J Hosp Med. 2009;4(9):E7-E14.

11. Matheny ME, Shubina M, Kimmel ZM, Pendergrass ML, Turchin A. Treatment Intensification and Blood Glucose Control among Hospitalized Diabetic Patients. J Gen Intern Med. 2008;23(2):184-189.

S.4. Numerator Statement: The total number of hyperglycemic days across all encounters divided by the total number of eligible days across all encounters. Hospital days are measured in 24-hour periods, starting from the time of arrival at the hospital (including Emergency Department). Days with a hyperglycemic event are defined as:

- A day with at least one blood glucose value >300 mg/dL; or

- A day in which a blood glucose value was not documented and it was preceded by two consecutive days where at least one glucose value is >=200 mg/dL.

We do not count >300 mg/DL events the first 24-hour period after admission to the hospital (including the Emergency Department) or the last time period before discharge, if it was less than 24 hours.

S.6. Denominator Statement: The initial population is all patients 18 years and older at the start of the measurement period with a discharged inpatient hospital admission during the measurement period, as well as either:

1. A diagnosis of diabetes that starts before or during the encounter; or

2. Administration of at least one dose of insulin or any anti-diabetic medication during the encounter; or

3. Presence of at least one blood glucose value >200 mg/dL at any time during the encounter.

The eCQM includes inpatient encounters which began in the Emergency Department or in observation status.

The denominator is the total number of eligible days across all encounters which match the initial population criteria. We do not count the the first 24-hour period after admission to the hospital (including the Emergency Department) or the last time period before the discharge, if it was less than 24

hours. By excluding the first 24 hours of admission, we allow for correction of severe hyperglycemia that was present on admission. By excluding the last time period before discharge if it was less than 24 hours, we account for the fact that hospitals may not always be able to check glucose during the last time period, especially if it is only a few hours long. Eligible encounters that exceed 10 days are truncated to equal 10 days.

S.8. Denominator Exclusions: N/A; 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? This eCQM is intended to be used simultaneously with the eCQM assessing severe hypoglycemia events (NQF#3503e) during acute short-term care admissions for patients age 18 years and older. The Hospital Harm - Severe Hypoglycemia eCQM functions as a balancing measure for this eCQM by minimizing the potential for unintended consequences such as fluctuating glycemic values and hypoglycemia.

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

HospitalHarm_HyperglycemiaEvidenceForm-637075210542799053.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.

Yes

1a. Evidence (subcriterion 1a)

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 3533e

Measure Title: Hospital Harm – Severe Hyperglycemia

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

Date of Submission: 11/1/2019

Instructions

- ***Complete 1a.1 and 1a.2 for all measures. If instrument-based measure, complete 1a.3.***
- ***Complete EITHER 1a.2, 1a.3 or 1a.4 as applicable for the type of measure and evidence.***
- ***For composite performance measures:***
  - ***A separate evidence form is required for each component measure unless several components were studied together.***
  - ***If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.***
- **All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.**
- **If you are unable to check a box, please highlight or shade the box for your response.**
- **Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.**

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

1a. Evidence to Support the Measure Focus

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

- **Outcome: [3] Empirical data demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service. If not available, wide variation in performance can be used as evidence, assuming the data are from a robust number of providers and results are not subject to systematic bias.**
- **Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.**
- **Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.**
- **Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.**
- **Efficiency: [6] evidence not required for the resource use component.**
- **For measures derived from patient reports, evidence should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.**
- **Process measures incorporating Appropriate Use Criteria: See NQF's guidance for evidence for measures, in general; guidance for measures specifically based on clinical practice guidelines apply as well.**

Notes

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

4. The preferred systems for grading the evidence are the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines and/or modified GRADE.

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

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

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

Outcome

☒ Outcome: Severe Hyperglycemia

☐ 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

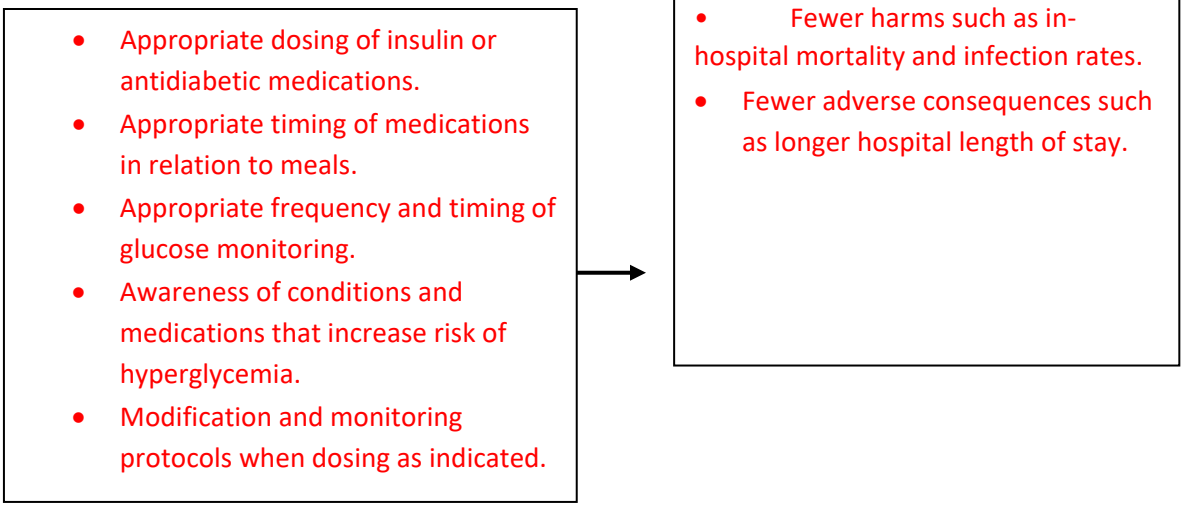
1a.2 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process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

The goal of this electronic clinical quality measure (eCQM) is to improve patient safety and prevent severe hyperglycemia in patients who are at higher risk. The focus of this outcome eCQM is inpatient severe hyperglycemia. The purpose of measuring severe hyperglycemic events is to reduce the frequency of these patient outcomes and to improve hospitals' practices for appropriate dosing of medication and adequate monitoring of patients with a diagnosis of diabetes mellitus, those receiving insulin or anti-diabetic medication(s), and those with an elevated glucose level during their hospital admission.

Severe hyperglycemia is significantly associated with a range of adverse consequences, including increased in-hospital mortality, infection rates, and hospital length of stay.[1,2,3,4,5,6,7,8] Additionally, hyperglycemia affects a wide range of inpatients, including individuals with no prior history of diabetes. Hyperglycemia may be induced in at-risk individuals by medications such as steroids, parenteral (intravenous) or enteral (tube) feeding, or critical illness.

The rate of inpatient severe hyperglycemia events can be considered a marker for quality of hospital care, since inpatient severe hyperglycemia is largely avoidable with proper glycemic management.

Studies indicate that use of evidence-based standardized protocols and insulin management protocols have been shown to improve glycemic control and safety outcomes.[9,10] Moreover, the rate of severe hyperglycemia varies across hospitals, suggesting opportunities for improvement in care. Hyperglycemic rates have been reported from 32.2%[11] to 46.0%[12] of ICU patient-days, and 31.7%[12] to 54.2%[13] of non-ICU patient-days (>180 mg/dL)



References:

1. Pasquel FJ, Spiegelman R, McCauley M, et al. Hyperglycemia During Total Parenteral Nutrition: An Important Marker of Poor Outcome and Mortality in Hospitalized Patients. Diabetes Care. 2010;33(4):739-741

2. Rady MY, Johnson DJ, Patel BM, Larson JS, Helmers RA. Influence of Individual Characteristics on Outcome of Glycemic Control in Intensive Care Unit Patients With or Without Diabetes Mellitus. Mayo Clin Proc. 2005;80(12):1558-1567.

3. Umpierrez GE, Isaacs SD, Bazargan N, You X, Thaler LM, Kitabchi AE. Hyperglycemia: An Independent Marker of In-Hospital Mortality in Patients with Undiagnosed Diabetes. J Clin Endocrinol Metab. 2002;87(3):978-982.

4. Falciglia M, Freyberg RW, Almenoff PL, D'Alessio DA, Render ML. Hyperglycemia-Related Mortality in Critically Ill Patients Varies with Admission Diagnosis. Crit Care Med. 2009;37(12):3001-3009.

5. Lee LJ, Emons MF, Martin SA, et al. Association of Blood Glucose Levels with In-Hospital Mortality and 30-Day Readmission in Patients Undergoing Invasive Cardiovascular Surgery. Curr Med Res Opin. 2012;28(10):1657-1665.

6. King JT, Jr., Goulet JL, Perkal MF, Rosenthal RA. Glycemic Control and Infections in Patients with Diabetes Undergoing Noncardiac Surgery. Ann Surg. 2011;253(1):158-165.

7. Jackson RS, Amdur RL, White JC, Macsata RA. Hyperglycemia is Associated with Increased Risk of Morbidity and Mortality after Colectomy for Cancer. J Am Coll Surg. 2012;214(1):68-80.

8. Umpierrez GE, Hellman R, Korytkowski MT, et al. Management of Hyperglycemia in Hospitalized Patients in Non-Critical Care Setting: An Endocrine Society Clinical Practice Guideline. J Clin Endocrinol Metab. 2012;97(1):16-38.

9. Maynard G, Kulasa K, Ramos P, et al. Impact of a Hypoglycemia Reduction Bundle and a Systems Approach to Inpatient Glycemic Management. Endocr Pract. 2015;21(4):355-367.

10. Donihi AC, DiNardo MM, DeVita MA, Korytkowski MT. Use of a Standardized Protocol to Decrease Medication Errors and Adverse Events Related to Sliding Scale Insulin. Qual Saf Health Care. 2006;15(2):89-91.

11. Swanson CM, Potter DJ, Kongable GL, Cook CB. Update on Inpatient Glycemic Control in Hospitals in the United States. Endocr Pract. 2011;17(6):853-861.

12. Cook CB, Kongable GL, Potter DJ, Abad VJ, Leija DE, Anderson M. Inpatient Glucose Control: A Glycemic Survey of 126 U.S. Hospitals. J Hosp Med. 2009;4(9):E7-E14.

13. Matheny ME, Shubina M, Kimmel ZM, Pendergrass ML, Turchin A. Treatment Intensification and Blood Glucose Control among Hospitalized Diabetic Patients. J Gen Intern Med. 2008;23(2):184-189.

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

N/A; this eCQM is not derived from patient report.

****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 use of evidence-based standardized protocols and insulin management protocols have been shown to improve glycemic control and safety outcomes.[14,15] This eCQM will improve inpatient glycemic control by promoting evidence-based interventions that optimize the care of patients with hyperglycemia and diabetes, which is one of 34 practices identified by NQF to reduce the occurrence of adverse healthcare events.[16] In the long term, the Hospital Harm - Severe Hyperglycemia eCQM provides a path to directly engage hospital staff and executives on the importance of glycemic measurement, and will be a tool for quality improvement for staff to assess internal metrics. In addition, this eCQM provides CMS an instrument to assess the quality of care to patients at risk for severe hyperglycemia across all acute care hospitals.

References:

14. Maynard G, Kulasa K, Ramos P, et al. Impact of a Hypoglycemia Reduction Bundle and a Systems Approach to Inpatient Glycemic Management. Endocr Pract. 2015;21(4):355-367.

15. Donihi AC, DiNardo MM, DeVita MA, Korytkowski MT. Use of a Standardized Protocol to Decrease Medication Errors and Adverse Events Related to Sliding Scale Insulin. Qual Saf Health Care. 2006;15(2):89-91.

16. National Quality Forum (NQF). Safe Practices for Better Healthcare–2010 Update: A Consensus Report. Washington, DC2010.

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 eCQM relates to glycemic management in the hospital inpatient setting. Rates of inpatient severe hyperglycemic events – an extremely elevated blood glucose level – can be considered an indicator of quality of care provided by a hospital. Severe hyperglycemia is associated with a range of harms, including increased in-hospital mortality, infection rates, and hospital length of stay.1,2,3,4,5,6,7,8 The rate of severe hyperglycemic events varies across hospitals, which suggests that there are opportunities for improvement in glycemic management.9,10,11 The implementation of this eCQM will aim to achieve several improvements in quality. For instance, this eCQM will encourage providers to develop interventions aimed at better glycemic control and prevent severe hyperglycemia for hospital inpatients. In addition to avoiding direct patient harm from the severe hyperglycemic event, lower rates of severe hyperglycemia among hospitalized individuals would be expected to result in lower rates of mortality, infection, and hospital length of stays. Adopting this eCQM has the potential to improve quality of care for individuals at risk of hyperglycemia and, therefore, advance the quality of care in patient safety, which is a priority area identified by the National Quality Strategy. This will fill a gap in measurement and provide incentives for hospital quality improvement, as there is no hyperglycemia measure in any CMS program. With a systematic EHR-based patient safety measure in place, hospitals can more reliably assess harm reduction efforts and modify their improvement efforts in near real-time. In addition, we can expect to make greater achievements in reducing harms and enhancing hospital performance on patient safety outcomes.

References:

1. Pasquel FJ, Spiegelman R, McCauley M, et al. Hyperglycemia During Total Parenteral Nutrition: An Important Marker of Poor Outcome and Mortality in Hospitalized Patients. Diabetes Care. 2010;33(4):739-741

2. Rady MY, Johnson DJ, Patel BM, Larson JS, Helmers RA. Influence of Individual Characteristics on Outcome of Glycemic Control in Intensive Care Unit Patients With or Without Diabetes Mellitus. Mayo Clin Proc. 2005;80(12):1558-1567.

3. Umpierrez GE, Isaacs SD, Bazargan N, You X, Thaler LM, Kitabchi AE. Hyperglycemia: An Independent Marker of In-Hospital Mortality in Patients with Undiagnosed Diabetes. J Clin Endocrinol Metab. 2002;87(3):978-982.

4. Falciglia M, Freyberg RW, Almenoff PL, D´Alessio DA, Render ML. Hyperglycemia-Related Mortality in Critically Ill Patients Varies with Admission Diagnosis. Crit Care Med. 2009;37(12):3001-3009.

5. Lee LJ, Emons MF, Martin SA, et al. Association of Blood Glucose Levels with In-Hospital Mortality and 30-Day Readmission in Patients Undergoing Invasive Cardiovascular Surgery. Curr Med Res Opin. 2012;28(10):1657-1665.

6. King JT, Jr., Goulet JL, Perkal MF, Rosenthal RA. Glycemic Control and Infections in Patients with Diabetes Undergoing Noncardiac Surgery. Ann Surg. 2011;253(1):158-165.

7. Jackson RS, Amdur RL, White JC, Macsata RA. Hyperglycemia is Associated with Increased Risk of Morbidity and Mortality after Colectomy for Cancer. J Am Coll Surg. 2012;214(1):68-80.

8. Umpierrez GE, Hellman R, Korytkowski MT, et al. Management of Hyperglycemia in Hospitalized Patients in Non-Critical Care Setting: An Endocrine Society Clinical Practice Guideline. J Clin Endocrinol Metab. 2012;97(1):16-38.

9. Swanson CM, Potter DJ, Kongable GL, Cook CB. Update on Inpatient Glycemic Control in Hospitals in the United States. Endocr Pract. 2011;17(6):853-861.

10. Cook CB, Kongable GL, Potter DJ, Abad VJ, Leija DE, Anderson M. Inpatient Glucose Control: A Glycemic Survey of 126 U.S. Hospitals. J Hosp Med. 2009;4(9):E7-E14.

11. Matheny ME, Shubina M, Kimmel ZM, Pendergrass ML, Turchin A. Treatment Intensification and Blood Glucose Control among Hospitalized Diabetic Patients. J Gen Intern Med. 2008;23(2):184-189.

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

This eCQM was tested with seven hospitals in four regions (West, Midwest, Southeast, South). Hospitals varied in size (100-799 beds) and EHR systems (Cerner, Meditech, Epic). Six of the seven hospitals were teaching hospitals, and two were located in rural areas while five hospitals were located in urban areas. A detailed list of the characteristics of measured facilities and patient population can be found in the attached Measure Testing Form, Section 1.7.

The measure performance, including the denominator observations (hospital days), numerator observations (hospital days), and performance rate by hospital, follows.

Hospital 1

- Data collection period: 1/1/2018 – 3/31/2018
- Denominator: 4,776
- Numerator: 510
- Performance rate: 10.1%
- 95% confidence interval: 9.8%, 11.6%
- Standard Deviation: 0.5%

Hospital 2

- Data collection period: 1/1/2018 – 3/31/2018
- Denominator: 1,362
- Numerator: 112
- Performance rate: 8.2%
- 95% confidence interval: 6.8%, 9.7%
- Standard Deviation: 0.7%

Hospital 3

- Data collection period: 1/1/2018 – 3/31/2018

- Denominator: 2,643
- Numerator: 330
- Performance rate: 12.5%
- 95% confidence interval: 11.2%, 13.7%
- Standard Deviation: 0.6%

Hospital 4

- Data collection period: 1/1/2018 – 3/31/2018
- Denominator: 4,219
- Numerator: 548
- Performance rate: 13.0%
- 95% confidence interval: 12.0%, 14.0%
- Standard Deviation: 0.5%

Hospital 5

- Data collection period: 1/1/2018 - 10/31/2018
- Denominator: 3,413
- Numerator: 667
- Performance rate: 19.5%
- 95% confidence interval: 18.2%, 20.9%
- Standard Deviation: 0.7%

Hospital 6

- Data collection period: 1/1/2018 - 10/31/2018
- Denominator: 3,323
- Numerator: 512
- Performance rate: 15.4%
- 95% confidence interval: 14.2%, 16.7%
- Standard Deviation: 0.6%

Hospital 7

- Data collection period: 1/1/2018 - 10/31/2018
- Denominator: 25,595
- Numerator: 3,865

Hospital 7 was not able to map POC glucose lab data at the time of testing, and therefore we could not include this hospital in the calculation of the performance rate. Of note, Hospital 7 did have POC glucose lab data that were available in structured fields, deemed accurate, and captured as part of normal clinical workflow, however these data were not codified using national standards nor were they mapped to such in the EHR system.

Overall Performance Rate for Hospitals 1-6

- Performance Rate: 13.6%

- 95% confidence interval: 13.1%, 14.1%

- Standard deviation: 0.2%

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, gender, race, and ethnicity.

Hospital 1 (Alpha dataset per Testing Form)

- Data collection period: 1/1/2018 – 3/31/2018

- Denominator (number of eligible hospital days): 4,776

Hospital 2 (Alpha dataset per Testing Form)

- Data collection period: 1/1/2018 – 3/31/2018

- Denominator (number of eligible hospital days): 1,362

Hospital 3 (Alpha dataset per Testing Form)

- Data collection period: 1/1/2018 – 3/31/2018

- Denominator (number of eligible hospital days): 2,643

Hospital 4 (Alpha dataset per Testing Form)

- Data collection period: 1/1/2018 – 3/31/2018

- Denominator (number of eligible hospital days): 4,219

Hospital 5 (Beta dataset per Testing Form)

- Data collection period: 1/1/2018 - 10/31/2018

- Denominator (number of eligible hospital days): 3,413

Hospital 6 (Beta dataset per Testing Form)

- Data collection period: 1/1/2018 - 10/31/2018

- Denominator (number of eligible hospital days): 3,323

Across Sites (n = 19,736, 6 hospitals)

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

18-64//7,471//1,128//15.1% (14.1%, 16.1%)

65+//12,265//1,551//12.6% (12.1%, 13.3%)

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

Male//10,037 //1,265//12.6% (12.0%, 13.3%)

Female//9,699//1,414//14.6% (13.9%, 15.3%)

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

Black or African American//1,597//219//13.7% (12.1%, 15.5%)

White//14,094//2,000//14.2% (13.6%, 14.8%)

Other//3,911//450//11.5% (10.5%, 12.6%)

Unknown//134//10//7.5% (3.6%, 13.3%)

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

Hispanic or Latino//1.149//126//11.0% (9.2%, 12.9%)

Non-Hispanic//12,876//1,887//14.7% (14.1%, 15.3%)

Unknown//5,711//666//11.7% (10.8%, 12.5%)

While testing meets or exceeds NQF requirements for number of facilities and EHRs, we note that these results are derived from a small dataset that may not be generalizable to the entire population. The disparities datasets include characteristics that may be documented as 'unknown' in some facilities, 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 reporting website, once the finalized through the NQF endorsement and CMS pre-rulemaking and rulemaking processes.

S.2a. <u>If this is an eMeasure</u>, 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:** HospitalHarm_BonnieTestCasesResults031519.pdf,HyperG_v5_7_Artifacts.zip,HospitalHarm_HyperglycemiaEvidenceForm.docx

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:** Hospital_Harm_Hyperglycemia_Feasibility_Scorecard.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. <u>For maintenance of endorsement:</u> 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. <u>For maintenance of endorsement,</u> please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.

Although this measure is not undergoing maintenance and is being considered for endorsement as a new measure, we would like to provide a comparison between this new eCQM and the previously NQF-endorsed measure from which it was adapted: #2362e Glycemic Control - Hyperglycemia, which was developed by the Health Services Advisory Group (HSAG). The predecessor measure has been subsequently retired by CMS. Adaptations made to the predecessor specifications are as follows:

Numerator differences:

- The current Hospital Harm – Severe Hyperglycemia eCQM defines a severe hyperglycemic event as a hospital day with at least one blood glucose value >300 mg/dL; or a day in which a blood glucose value was not documented, and it was preceded by two consecutive days where at least one glucose value is >=200 mg/dL.

- The previous NQF-endorsed measure defined hyperglycemic hospital days as days in which: (1) two or more blood glucose levels were elevated (>200 mg/dL [11.1 mmol/L]), measured at least six hours apart; Or (2) a single blood glucose level was elevated, if only one value was available that day; Or (3) no blood glucose level was measured that day, and it was not preceded by two normoglycemic days.

Rationale for the change: Clinical experts supported using a higher threshold to define severe hyperglycemia (>300 mg/dL) as a clearer indication of patient harm. The higher threshold will likely improve the acceptability among clinicians and avoid the unintended consequence of hypoglycemia.

Denominator differences: N/A

Exclusion differences:

- The current Hospital Harm – Severe Hyperglycemia eCQM includes patients on metformin and patients admitted with a diabetic ketoacidosis (DKA) or hyperosmolar hyperglycemic state (HHS).

- The previous NQF-endorsed measure excluded patients on metformin only (i.e., for polycystic ovarian syndrome) and patients admitted with a diagnosis of DKA or HHS.

Rationale for the change: In testing for the current eCQM, a negligible number of patients entered the denominator only through the use of metformin. Clinical experts advised that hospitals should be able to decrease glucose levels below 300 mg/dL within 24 hours for patients admitted for DKA or HHS.

Definition of Hospital Days:

- The current Hospital Harm – Severe Hyperglycemia eCQM uses 24-hour windows starting with the arrival date and time to define hospital days.

- The previous NQF-endorsed measure used calendar days to define hospital days.

Rationale for the change: The new approach is easier for hospitals to compute and simplifies the eCQM logic to exclude hyperglycemia that is present on admission.

Risk Adjustment/Stratification:

- The current Hospital Harm – Severe Hyperglycemia eCQM is not risk adjusted or stratified.

- The previous NQF-endorsed measure stratified results by care units (intensive care unit [ICU] vs. non-ICU), type of patients (medical vs. surgical), and daily cumulative steroid dose (<10 mg, 10-499 mg, and >500 mg prednisone equivalents).

Rationale for the change: Input from our clinical experts indicated that although patients in the ICU and those receiving steroids are at increased risk of hyperglycemia, extreme values over 300 mg/dL are avoidable with careful monitoring and proper medical management.

Measure Calculation:

- The current Hospital Harm – Severe Hyperglycemia eCQM calculates the total number of hyperglycemic days across all encounters divided by the total number of eligible days across all encounters.

- The previous NQF-endorsed measure calculated the average percentage of hyperglycemic hospital days in hyperglycemia for each admission.

Rationale for the change: The new approach accounts for the length of stay and mitigates the impact of extreme values on a hospital's score.

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 total number of hyperglycemic days across all encounters divided by the total number of eligible days across all encounters. Hospital days are measured in 24-hour periods, starting from the time of arrival at the hospital (including Emergency Department). Days with a hyperglycemic event are defined as:

- A day with at least one blood glucose value >300 mg/dL; or

- A day in which a blood glucose value was not documented and it was preceded by two consecutive days where at least one glucose value is >=200 mg/dL.

We do not count >300 mg/DL events the first 24-hour period after admission to the hospital (including the Emergency Department) or the last time period before discharge, if it was less than 24 hours.

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 (EHR) data to calculate the measure score. The 24-hour window for data collection is during an inpatient hospitalization, beginning at hospital arrival (whether through the Emergency Department, observation stay, or direct admission to inpatient).

All data elements necessary to calculate this eCQM are defined within value sets available in the Value Set Authority Center (VSAC) and listed below.

Glucose tests are represented by LOINC codes in the value set Glucose Lab Test (2.16.840.1.113762.1.4.1045.134). Codes include laboratory and point-of-care glucose tests, including glucose in blood, serum or plasma, venous blood, and arterial blood; and fasting glucose in venous blood and serum or plasma.

To access the value sets for the eCQM, 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)*

The initial population is all patients 18 years and older at the start of the measurement period with a discharged inpatient hospital admission during the measurement period, as well as either:

1. A diagnosis of diabetes that starts before or during the encounter; or

2. Administration of at least one dose of insulin or any anti-diabetic medication during the encounter; or

3. Presence of at least one blood glucose value >200 mg/dL at any time during the encounter.

The eCQM includes inpatient encounters which began in the Emergency Department or in observation status.

The denominator is the total number of eligible days across all encounters which match the initial population criteria. We do not count the the first 24-hour period after admission to the hospital (including the Emergency Department) or the last time period before the discharge, if it was less than 24 hours. By excluding the first 24 hours of admission, we allow for correction of severe hyperglycemia that was present on admission. By excluding the last time period before discharge if it was less than 24 hours, we account for the fact that hospitals may not always be able to check glucose during the last time period, especially if it is only a few hours long. Eligible encounters that exceed 10 days are truncated to equal 10 days.

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 eCQM includes all patients 18 years and older at the start of the measurement period, and all payers. The measurement period is 12 months.

- Glucose tests are represented by LOINC codes in the value set Glucose Lab Test (2.16.840.1.113762.1.4.1045.134).

- Inpatient Encounters are represented using the value set of SNOMEDCT codes (2.16.840.1.113883.3.666.5.307).

- Emergency Department Visits are represented using the value set of SNOMEDCT codes (2.16.840.1.113883.3.117.1.7.1.292).

- Observation Services are represented using the value set of SNOMEDCT codes (2.16.840.1.113762.1.4.1111.143).

- Patients who were given at least one administration of insulin or any anti-diabetic medication during the encounter are defined by the value set of RXNORM codes (2.16.840.1.113883.3.1260.1.1978). This value set includes medications and insulin capable of causing severe hyperglycemia (blood glucose value >300 mg/dL).

- Diabetes are represented using the value set of ICD10CM, ICD9CM, SNOMEDCT codes (2.16.840.1.113883.3.464.1003.103.12.1001). This value set includes patients diagnosed with diabetes before or during the encounter.

To access the value sets for the eCQM, please visit the Value Set Authority Center, 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)*

N/A; 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

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 eCQM 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:

Ratio

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 encounters, all payers, where individuals are aged 18 years and older at the start of the measurement period and have:

1. A diagnosis of diabetes that starts before or during the encounter; or

2. Administration of at least one dose of insulin or any anti-diabetic medication during the encounter; or

3. Presence of at least one blood glucose value >200 mg/dL at any time during the encounter.

To create the denominator:

1. If the inpatient encounter occurred during the measurement period, go to Step 2. If not, do not include in the denominator.

2. Determine the patient's age in years. The patient's age is equal to the measurement period start date minus the birth date. If the patient is at least 18 years old, go to Step 3. If less than 18 years old, do not include in the denominator.

3. Determine if the patient had a diagnosis of diabetes mellitus before or during the hospital encounter, or if the patient was administered at least one dose of insulin or an anti-diabetic medication during the encounter, or if the patient had a glucose level of >200 mg/dL during the hospital encounter. If any of these three conditions exist, then include in the denominator. If not, do not include in the denominator.

4. (As the denominator is measured in days, which are defined as 24-hour periods starting at the time of arrival to the hospital (including the Emergency Department)): if the 24-hour period is not the first 24-hour period of the hospital admission, and is not the last period prior to hospital discharge if less than 24 hours, then include in the denominator. If it is the first 24-hour period or the last period prior to discharge that is less than 24 hours, do not include in the denominator.

a) By excluding for >300 mg/dL events the first 24 hours of admission, we allow for correction of severe hyperglycemia that was present on admission. By excluding the last time period before discharge if it was less than 24 hours, we account for the fact that hospitals may not always be able to check glucose during the last time period, especially if it is only a few hours long.

To create the numerator:

1. During any 24-hour period from arrival to the hospital (including the Emergency Department) except for the first 24-hour period and the last period prior to hospital discharge if less than 24 hours, any 24-hour period with a blood glucose level >300 mg/dL;

Or

2. A 24-hour period in which a blood glucose value was not documented, and it was preceded by two consecutive days where at least one glucose value is >=200 mg/dL.

If either of these 2 events occur, then include in the numerator. If not, do not include in the numerator.

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 eCQM does not use a sample or survey.

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 eCQM does not use a sample or 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 measure authoring tool (MAT) output, which includes the human readable and XML artifacts of the clinical quality language (CQL) for the eCQM are contained in the 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

Hospital_Harm_Hyperglycemia_Testing_Attachment_7.29.19.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*)**:** Click here to enter NQF number
Measure Title: Hospital Harm – Severe Hyperglycemia
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:	Measure Tested with Data From:

(*must be consistent with data sources entered in S.17*)	
☐ 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 for four hospitals to test the measure concept: feasibility, validity, and reliability. We additionally partnered with three hospitals to complete beta testing of the Measure Authoring Tool (MAT) output in two different electronic health record (EHR) systems. Using these data we assessed measure score reliability and data element validity as well as missing data. 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 <u>all</u> the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.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 hospitals, dates, and number of admissions used in each phase of testing are in **Table 1**.

Table 1. Dataset Descriptions

Hospital	Applicable Section in the Testing Attachment	Description of Dataset	EHR Vendor	Phase
Hospital 1	Section 2a2 Reliability Testing Section 2b1 Validity Testing (Measure Score) Section 2b4 Identification of Statistically Significant and Meaningful Differences in Performance	Data obtained from Patient Safety Organization Dates of Data: January 1, 2018 - March 31, 2018 Number of Hospital Days: 4,776 Number of Unique Patients: 1,325 For Validity Testing: sample of 200 admissions This is an urban, teaching hospital with a bed size of 200-299 beds. Located in the West.	Cerner	Alpha
Hospital 2	Section 2a2 Reliability Testing Section 2b1 Validity Testing (Measure Score)	Data obtained from Patient Safety Organization Dates of Data: January 1, 2018 - March 31, 2018	Cerner	Alpha

	Section 2b4 Identification of Statistically Significant and Meaningful Differences in Performance	Number of Hospital Days: 1,362 Number of Unique Patients: 469 For Validity Testing: sample of 200 admissions This is an urban, teaching hospital with a bed size of 100-199 beds. Located in the South.		
Hospital 3	Section 2a2 Reliability Testing Section 2b1 Validity Testing (Measure Score) Section 2b4 Identification of Statistically Significant and Meaningful Differences in Performance	Data obtained from Patient Safety Organization Dates of Data: January 1, 2018 - March 31, 2018 Number of Hospital Days: 2,643 Number of Unique Patients: 935 For Validity Testing: sample of 200 admissions This is an urban, teaching hospital with a bed size of 200-299 beds. Located in the West.	Cerner	Alpha
Hospital 4	Section 2a2 Reliability Testing Section 2b1 Validity Testing (Measure Score) Section 2b4 Identification of Statistically Significant and Meaningful Differences in Performance	Data obtained from Patient Safety Organization Dates of Data: January 1, 2018 - March 31, 2018 Number of Hospital Days: 4,219 Number of Unique Patients: 1,241 For Validity Testing: sample of 200 admissions	Epic	Alpha

		This is an urban, non-teaching hospital with a bed size of 300-399 beds. Located in the West.		
Hospital 5	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, 2018 - October 31, 2018 Number of Hospital Days: 3,413 Number of Unique Patients: 868 For Validity Testing: sample of 175 hospital days (data element) and 100 numerator hospital days (measure score) For Missing Data Analysis: sample of 175 hospital days This is a rural, non-teaching hospital with a bed size of 100-199 beds. Located in the Midwest.	Meditech	Beta
Hospital 6	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, 2018 - October 31, 2018 Number of Hospital Days: 3,323 Number of Unique Patients: 663 For Validity Testing: sample of 175 hospital days (data element) and 100 numerator hospital days (measure score) For Missing Data Analysis: sample of 175 hospital days	Meditech	Beta

		This is a rural, teaching hospital with a bed size of 200-299 beds. Located in the Midwest.		
Hospital 7	Section 2b1 Validity Testing (Data Element) Section 2b6 Missing Data Analysis	Dates of Data: January 1, 2018 - October 31, 2018 Number of Hospital Days: 25,595 Number of Unique Patients: 4,337 For Validity Testing: sample of 175 hospital days For Missing Data Analysis: sample of 175 hospital days This is an urban, teaching hospital with a bed size of 700-799 beds. Located in the South.	Epic	Beta

Hospital 7 was not able to map POC glucose lab data, and therefore we could not include this hospital in the calculation of the measure score reliability, measure score validity (PPV), and performance rate. When possible, testing would ideally include engagement with the vendor to help support system needs for measure implementation. In this instance, the local and/or vendor codes were not already mapped to those in the measure value set specific to glucose values. Since the vendor was not engaged in the testing process, and therefore did not complete the mapping, data would have been incomplete in terms of calculating measure score reliability, validity and performance rates.

The adjudication process was able to appropriately identify the presence of lab values in the EHRs. However, that data were not retrievable in a report by the test site, which was set to run based upon the defined codes from the value set. With incentive from CMS (by adding to a rule and requiring implementation), the mapping would be completed by EHRs vendors in advance and would thus enable full implementation at the organization in question.

Patient descriptive characteristics included in the analysis by hospital for **Hospitals 1-7** are provided below:

Initial Patient Population Characteristics	Hospital 1		Hospital 2		Hospital 3		Hospital 4	
	n	%	n	%	n	%	n	%

Number of unique patients	1,325	100.00%	469	100.00%	935	100.00%	1,241	100.00%
Average Age [Mean (STD)]	66 (14)		69 (14)		66 (15)		68 (15)	
18-35	37	2.79%	9	1.92%	33	3.53%	52	4.19%
36-64	561	42.34%	160	34.12%	411	43.96%	417	33.60%
65+	727	54.87%	300	63.97%	491	52.51%	772	62.21%
Sex								
Male	702	52.98%	265	56.50%	414	44.28%	627	50.52%
Female	623	47.02%	204	43.50%	521	55.72%	614	49.48%
Race								
Black or African American	194	14.64%	29	6.18%	194	20.75%	58	4.67%
White	851	64.23%	417	88.91%	385	41.18%	574	46.25%
Other	269	20.30%	22	4.69%	352	37.65%	593	47.78%
Unknown	11	0.83%	1	0.21%	4	0.43%	16	1.29%
Ethnicity								
Hispanic or Latino	196	14.79%	-	0.00%	163	17.43%	-	0.00%
Non-Hispanic	1,099	82.94%	-	0.00%	762	81.50%	-	0.00%
Unknown	30	2.26%	469	100.00%	10	1.07%	1,241	100.00%
(Primary) Payer	-	-	-	-	-	-	-	-
Medicare	-	-	-	-	-	-	-	-
Medicaid	-	-	-	-	-	-	-	-
Private Insurance	-	-	-	-	-	-	-	-
Self-pay or Uninsured	-	-	-	-	-	-	-	-
Other	-	-	-	-	-	-	-	-
Unknown	-	-	-	-	-	-	-	-

Initial Patient Population Characteristics	**Hospital 5**		**Hospital 6**		**Hospital 7**		**Across Hospitals**	
	n	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Number of unique patients	868	100.00%	663	100.00%	4,337	100.00%	9,838	100.00%
Average Age [Mean (STD)]	68 (15)		69 (14)		58 (16)		63 (16)	
18-35	37	4.26%	16	2.41%	469	10.81%	653	10.01%
36-64	283	32.60%	212	31.98%	2,146	49.48%	4,190	64.26%
65+	548	63.13%	435	65.61%	1,722	39.70%	4,995	76.61%
Sex								
Male	427	49.19%	325	49.02%	2,218	51.14%	4,978	50.60%
Female	441	50.81%	338	50.98%	2,119	48.86%	4,860	49.40%
Race								
Black or African American	22	2.53%	22	3.32%	1,811	41.76%	2,330	23.68%
White	837	96.43%	637	96.08%	2,342	54.00%	6,043	61.43%
Other	6	0.69%	3	0.45%	168	3.87%	1,413	14.36%
Unknown	3	0.35%	1	0.15%	16	0.37%	52	0.53%

Ethnicity								
Hispanic or Latino	2	0.23%	2	0.30%	107	2.47%	470	4.78%
Non-Hispanic	864	99.54%	661	99.70%	4,217	97.23%	7,603	77.28%
Unknown	2	0.23%	-	0.00%	13	0.30%	1,765	17.94%
(Primary) Payer								
Medicare	636	73.27%	499	75.26%	2,379	54.85%	3,514	35.72%
Medicaid	122	14.06%	96	14.48%	579	13.35%	797	8.10%
Private Insurance	91	10.48%	52	7.84%	992	22.87%	1,135	11.54%
Self-pay or Uninsured	14	1.61%	-	0.00%	189	4.36%	203	2.06%
Other	4	0.46%	12	1.81%	198	4.57%	214	2.18%
Unknown	1	0.12%	4	0.60%	-	0.00%	5	0.05%

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, we collected information on the following social risk factors using data extracted from hospital EHR systems: race, ethnicity, and primary payer (if available).

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*)

Reliability testing demonstrates that 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.

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 **Hospitals 1-6**.

We assessed signal-to-noise reliability that describes how well the measure can distinguish the performance of one hospital from another.[1,2] The signal is the proportion of the variability in measured performance that can be explained by real differences in performance. Scores can range from zero to one. 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 to calculate the signal-to-noise ratio reliability.[3] 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 p̂ is the harm rate of a hospital.

References:

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

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

3. 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 5,501 eligible encounters (and 19,736 eligible days) across Hospitals 1-6. The signal-to-noise ratio yielded a median reliability score of 0.967 (range: 0.955-0.983).

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 0.967 indicates excellent agreement.

Our interpretation of these results is based on the standards established in literature.[4]

< 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

References:

4. 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 performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*) **NOTE**: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

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

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

Data Element Validity

For **Hospitals 5-7**, a stratified sample of 175 total discharges (stratifying by numerator encounters and denominator-only encounters) was 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 the positive predictive value (PPV), which is applicable for both data element validity and measure score validity. We adjudicated all numerator cases in alpha testing (in Hospitals 1 – 4) 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 one hospital would give us an even more accurate estimation.

Hospitals 5-7 each had 175 encounters, 100 encounters with at least one harm event (numerator hospital day) and 75 encounters with no harm events (denominator-only). Data were abstracted from the EHR by trained abstractors. Abstractors were provided with an excel spreadsheet to document the information abstracted from the EHR.

Table 2 shows the sensitivity agreement rate (# exact matches in both data sources / # sampled in the chart) between the data extracted from the EHR electronically and manual chart abstraction in **Hospitals 5-7**. Each data element matched if the electronically extracted value exactly matched the manually abstracted value (gold standard). For data/time data elements, we matched year, month, day, hour, and minutes. For glucose lab values, we matched on the glucose value result (whole integers), date, and time

within one minute. For administration of antidiabetic medications, we matched on the name of the medication administered.

Empirical Measure Score Validity

Measure score validity assesses whether the harm rate (or the measure score outcome) calculated for each facility is accurate. The measure score is calculated for each facility based on the number of hospital days across all encounters that experienced a harm compared to the total number of encounter days. 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 occurs 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 according to the medical record 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 a day that met the criteria for a harm among the sample of abstracted records. We clinically adjudicated 200 encounters for **Hospitals 1-4**, and 100 numerator hospital days for **Hospitals 5 and 6**. We then calculated the (PPV) for all numerator hospital days for **Hospitals 1-6**, as shown in **Table 3**. The PPV describes the probability that a patient with a positive result (numerator day) in the EHR data also had a positive result (numerator day) in the abstracted medical record data, as confirmed by a clinical adjudicator.

PPV: true positive / (true positive + false positive)

We also calculated the sensitivity, specificity, kappa, and negative predictive value (NPV) as shown in **Table 4** for **Hospitals 1-4**. Sensitivity describes the probability that a patient with a positive result in the abstracted medical record data also had 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 had a negative result in the abstracted medical record, confirmed by the clinical adjudicator.

Face Validity:

To systematically assess face validity, we surveyed our Technical Expert Panel (TEP), which is comprised of national experts and stakeholder organizations. We asked each member to rate the following statement using a six-point scale (1=Strongly Disagree, 2=Moderately Disagree, 3=Somewhat Disagree, 4=Somewhat Agree, 5= Moderately Agree, and 6=Strongly Agree): “the proportion of severe hyperglycemic events obtained from the Hospital Harm – Severe Hyperglycemia Electronic Clinical Quality Measure (eCQM) as specified can be used to distinguish between better and worse quality care at hospitals.

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

Data Element Validity

Table 2. Data Element Validity (Sensitivity) Results Required for Measure (Hospitals 5-7)

Data Element	Hospital 5			Hospital 6			Hospital 7		
	# Cases Matched in EHR (n)	# Cases in Abstraction (n)	Sensitivity Percent Match (%)	# Cases Matched in EHR (n)	# Cases in Abstraction (n)	Sensitivity Percent Match (%)	# Cases Matched in EHR (n)	# Cases in Abstraction (n)	Sensitivity Percent Match (%)
Admission date and time (mm/dd/yyyy, hh:mm)	175	175	100.0%	175	175	100.0%	175	175	100.0%
Discharge date and time (mm/dd/yyyy, hh:mm)	175	175	100.0%	175	175	100.0%	175	175	100.0%
Diabetes diagnosis	144	144	100.0%	129	129	100.0%	131	131	100.0%
Medication administered, antidiabetic medication name	125	125	100.0%	80	80	100.0%	151	151	100.0%
Laboratory test and point-of-care blood glucose results with date and time (mm/dd/yyy hh:mm result)	175	175	100.0%	173	173	100.0%	168	168	100.0%

Empirical Measure Score Validity

Table 3 displays the PPV for **Hospitals 1-6**. This PPV represents the percent of encounters that met the criteria for a harm (numerator) in the EHR confirmed by the chart abstraction, validated by an adjudicator. **Table 4** displays the specificity, sensitivity, kappa, and NPV for **Hospitals 1-4**.

Table 3. Measure Score Validity Statistics for Sample Between Electronic EHR Extraction and Manual Chart Abstraction (Hospitals 1-6)

Measure Component	Hospital 1 PPV	Hospital 2 PPV	Hospital 3 PPV	Hospital 4 PPV	Hospital 5 PPV	Hospital 6 PPV
Numerator	87.7%	100.0%	95.7%	98.6%	99.0%	100.0%

Table 4. Measure Score Validity Statistics for Sample Between Electronic EHR Extraction and Manual Chart Abstraction (Sensitivity, Specificity, NPV, Kappa) (Alpha testing Hospitals 1-4)

Measure	Hospital 1 (N= 1,346)				Hospital 2 (N=1,057)				Hospital 3 (N=1,262)				Hospital 4 (N=1,313)			
	Sensitivity	Specificity	Kappa (95% CI)	NPV	Sensitivity	Specificity	Kappa (95% CI)	NPV	Sensitivity	Specificity	Kappa (95% CI)	NPV	Sensitivity	Specificity	Kappa (95% CI)	NPV
Severe Hyperglycemia	100%	98.9%	0.92 (0.89, 0.96)	100%	100%	100%	1 (1,1)	100%	90.3%	99.7%	0.93 (0.89, 0.96)	99.0%	100%	99.7%	0.98 (0.97,1)	99.9%

Face Validity

10 out of 11 TEP members responded to the face validity survey question posed for this measure and answered as follows: Strongly Disagreed (0), Moderately Disagreed (0), Somewhat Disagreed (0), Somewhat Agreed (2), Moderately Agreed (4), and Strongly Agreed (4). Many TEP members commented that the measure was important, especially if reported in concert with a measure of hypoglycemia.

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 data elements had a match rate of 100%, indicating that valid and accurate data elements were extracted from the EHR. For the blood glucose date, time, and result data element, we assessed the validity of all glucose values recorded during the hospitalization for a more robust sample to evaluate a clearer picture of data element accuracy. Overall, the data elements required for the eCQM show validity.

Empirical Measure Score Validity

All six hospitals had a PPV over 87%, five with PPVs above 95%, indicating that in almost all cases the encounter met the criteria for a harm in both the chart abstracted and EHR-extracted data. Although we do not always expect perfect agreement, as we expect some degree of human error in entering and matching values, we consider the PPV to show excellent measure score validity. The absence of a perfect PPV does not threaten validity as we do not expect any systematic error in this small amount of disagreement across hospitals that might bias the measure results. Similarly, specificity and sensitivity are high. Sensitivity is 100% in **Hospitals 1, 2, and 4** and 90.3% in **Hospital 3**. Specificity is 100% in **Hospital 2**, 98.9% in **Hospital 1**, and 99.7% in **Hospitals 3 and 4**. This means that the probability of the EHR data detecting a true severe hyperglycemic event in patients that had a true severe hyperglycemic event based on the abstracted data ('gold standard') is 90-100% (sensitivity). The probability of the EHR data detecting no hyperglycemia when no hyperglycemic event occurred based on abstracted data is 99-100% (specificity). NPV was 100% in **Hospitals 1 and 2**, 99.0% in **Hospital 3**, and 99.9% in **Hospital 4**, indicating that the EHR data indicated that a harm did not occur, and 99-100% of the time the chart abstraction confirmed a harm did not occur. Kappa of 0.92, 1, and 0.93, and 0.98 indicate almost perfect agreement.[5]

Our Kappa interpretation is based on the following standards set in the literature:[6]

0.4 – 0.6 indicate “moderate agreement”,

0.6 – 0.8 “substantial agreement”, and

0.8 – 1 “almost perfect agreement”

Face Validity:

100% of TEP members agreed (somewhat, moderately, or strongly) that the eCQM will provide an accurate reflection of quality, which reflects good face validity.

References:

5. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960;20:37–46.

6. 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 not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

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 degrees of risk, it is important to consider risk adjustment for patient risk factors to compare hospital performance.

However, many harms such as severe hyperglycemia are avoidable, regardless of patient risk. We consider the following criteria in determining whether risk adjustment is warranted for the Severe Hyperglycemia eCQM:

1. If many patients are at risk of the harm regardless of their age, clinical status, comorbidities, or reason for admission, as described further in paragraph below;

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 or inadequate monitoring; and

3. If there is evidence that the risk of a harm can be largely ameliorated by best care practices regardless of a patient's 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 the Severe Hyperglycemia eCQM, there is evidence indicating that most hyperglycemic events of this severity (>300 mg/dL) are avoidable. There are several factors that affect glucose levels, including medications such as steroids, critical illness, infection, and other factors such as type 1 (versus type 2) diabetes. However, physicians should be able to achieve glucose levels of <300 mg/dL in all these cases, although the strategies to achieve them may differ depending on the circumstances. As these causes are controllable in hospital environments, and risk can easily be reduced by following best practices, we do not think risk adjustment is warranted for this eCQM. We will continue to evaluate the appropriateness of risk adjustment in measure reevaluation as is required for NQF endorsement maintenance.

In addition to the clinical rationale provided for not risk adjusting this eCQM, we examined the performance (harm) rate of the measure across patient characteristics of age, sex, race, ethnicity, and payer type (if available). 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.

Performance rate by encounter characteristic for **Hospitals 1-6** are provided below. Please note that payer information was not captured for alpha hospitals (1-4); for this reason, summary statistics across hospitals for payer type were not calculated (as they would only encompass two hospitals).

Characteristic	Hospital 1				
	Denominator	**Numerator**	**Performance Rate %**	**95% Confidence Interval**	**Standard deviation**
Number of Hospital Days	4,776	510	10.1%	9.8%-11.6%	0.5%
Age					
18-35	78	29	37.2%	26.5%-48.9%	5.5%
36-64	2,008	212	10.6%	9.3%-11.2%	0.7%
65+	2,690	269	10.0%	8.9%-11.2%	0.6%
Gender					

Male	2,653	242	9.1%	8.1%-10.3%	0.6%
Female	2,123	268	12.6%	11.2%-14.1%	0.7%
Race					
Black or African American	603	62	10.3%	8.0%-13.0%	1.2%
White	3,159	349	11.0%	10.0%-12.2%	0.6%
Other	965	97	10.1%	8.2%-12.1%	1.0%
Unknown	49	2	4.1%	0.5%-14.0%	2.8%
Ethnicity					
Hispanic or Latino	630	66	10.5%	8.2%-13.1%	1.2%
Non-Hispanic	4,038	439	10.9%	9.9%-11.9%	0.5%
Unknown / Unmapped	108	5	4.6%	1.5%-10.5%	2.0%

Characteristic	**Hospital 2**				
	Denominator	**Numerator**	**Performance Rate %**	**95% Confidence Interval**	**Standard deviation**
Number of Hospital Days	1,362	112	8.2%	6.8%-9.7%	0.7%
Age					
18-35	14	-	0.0%	0.0%-23.2%	0.0%
36-64	428	37	8.6%	6.2%-11.7%	1.4%
65+	920	75	8.2%	6.5%-10.1%	0.9%
Gender					
Male	708	46	6.5%	4.8%-8.6%	0.9%
Female	654	66	10.1%	7.9%-12.7%	1.2%
Race					
Black or African American	106	11	10.4%	5.3%-17.8%	3.0%
White	1,220	93	7.6%	6.0%-9.3%	0.8%
Other	34	8	23.5%	10.8%-41.2%	7.3%
Unknown	2	-	0.0%	0.0%-84.2%	0.0%
Ethnicity					
Hispanic or Latino	-	-			

Non-Hispanic	-	-			
Unknown / Unmapped	1,362	112	8.2%	6.8%-9.8%	0.7%

Characteristic	Hospital 3				
	Denominator	**Numerator**	**Performance Rate %**	**95% Confidence Interval**	**Standard deviation**
Number of Hospital Days	2,643	330	12.5%	11.2%-13.7%	0.60%
Age					
18-35	73	13	17.8%	9.8%-28.5%	4.5%
36-64	1,115	141	12.6%	10.8%-14.7%	1.0%
65+	1,455	176	12.1%	10.5%-13.9%	0.9%
Gender					
Male	1,224	134	10.9%	9.3%-12.8%	0.9%
Female	1,419	196	13.8%	12.1%-15.7%	0.9%
Race					
Black or African American	517	77	14.9%	11.9%-18.3%	1.6%
White	1,166	135	11.6%	9.8%-13.6%	0.9%
Other	948	118	12.4%	10.4%-14.7%	1.1%
Unknown	12	-	0.0%	0.0%-26.5%	0.0%
Ethnicity					
Hispanic or Latino	508	59	11.6%	9.0%-14.7%	1.4%
Non-Hispanic	2,115	271	12.8%	11.4%-14.3%	0.7%
Unknown / Unmapped	20	-	0.0%	0.0%-16.8%	0.0%

Characteristic	Hospital 4				
	Denominator	**Numerator**	**Performance Rate %**	**95% Confidence Interval**	**Standard deviation**
Number of Hospital Days	4,219	548	13.0%	12.0%-14.0%	0.5%
Age					

18-35	128	23	18.0%	11.7%-25.7%	3.4%
36-64	1,292	204	15.8%	13.8%-17.9%	1.0%
65+	2,799	321	11.5%	10.3%-12.7%	0.6%
Gender					
Male	2,112	296	14.0%	12.6%-15.6%	0.8%
Female	2,107	252	12.0%	10.6%-13.4%	0.7%
Race					
Black or African American	236	35	14.8%	10.6%-20.0%	2.3%
White	1,992	278	14.0%	12.5%-15.6%	0.8%
Other	1,926	227	11.8%	10.4%-13.3%	0.7%
Unknown	65	8	12.3%	5.5%-22.8%	4.1%
Ethnicity					
Hispanic or Latino	-	-			
Non-Hispanic	-	-			
Unknown / Unmapped	4,219	548	13.0%	12.0%-14.1%	0.5%

Characteristic	**Hospital 5**			
	Denominator	**Numerator**	**Performance Rate %**	**95% Confidence Interval**
Number of Hospital Days	3,413	667	19.5%	18.2%-20.9%
Age				
18-35	119	30	25.2%	17.7%-34.0%
36-64	1,151	256	22.2%	19.9%-24.8%
65+	2,143	381	17.8%	16.2%-19.5%
Gender				
Male	1,692	316	18.7%	16.9%-20.6%
Female	1,721	351	20.4%	18.5%-22.4%
Race				
Black or African American	54	21	38.9%	26.0%-53.1%
White	3,333	646	19.4%	18.1%-20.8%
Other	22	-	0.0%	0.0%-15.4%

Unknown	4	-	0.0%	0.0%-60.2%
Ethnicity				
Hispanic or Latino	6	1	16.7%	0.4%-64.1%
Non-Hispanic	3,405	665	19.5%	18.2%-20.1%
Unknown / Unmapped	2	1	50.0%	1.3%-98.7%
(Primary) Payer				
Medicare	2,538	490	19.3%	
Medicaid	520	119	22.9%	
Private Insurance	304	50	16.4%	
Self-pay or Uninsured	40	6	15.0%	
Other	11	2	18.2%	
Unknown	-	-	-	

Characteristic	**Hospital 6**			
	Denominator	**Numerator**	**Performance Rate %**	**95% Confidence Interval**
Number of Hospital Days	3,323	512	15.4%	14.2%-16.7%
Age				
18-35	39	22	56.4%	39.6%-72.2%
36-64	1,026	161	15.7%	13.5%-18.1%
65+	2,258	329	14.6%	13.1%-16.1%
Gender				
Male	1,648	231	14.0%	12.4%-15.8%
Female	1,675	281	16.8%	15.0%-18.7%
Race				
Black or African American	81	13	16.0%	8.9%-25.9%
White	3,224	499	15.5%	14.3%-16.8%
Other	16	-	0.0%	0.0%-20.6%
Unknown	2	-	0.0%	0.0%-84.2%
Ethnicity				
Hispanic or Latino	5	-	0.0%	0.0%-52.2%

Non-Hispanic	3,318	512	15.4%	14.2%-16.7%
Unknown / Unmapped	-	-	-	-
(Primary) Payer				
Medicare	2,629	378	14.4%	
Medicaid	459	87	19.0%	
Private Insurance	162	21	13.0%	
Self-pay or Uninsured	-	-	-	
Other	54	22	40.7%	
Unknown	19	4	21.1%	

Characteristic	Across Hospitals				
	Denominator	**Numerator**	**Performance Rate %**	**95% Confidence Interval**	**Range**
Number of Hospital Days	19,736	2,679	13.6%	13.1%-14.1%	8.2%-19.5%
Age					
18-35	451	117	25.9%	22.0%-30.3%	0.0%-56.4%
36-64	7,020	1,011	14.4%	13.6%-15.2%	8.6%-22.2%
65+	12,265	1,551	12.6%	12.1%-13.3%	8.2%-17.8%
Gender					
Male	10,037	1,265	12.6%	12.0%-13.3%	6.5%-18.7%
Female	9,699	1,414	14.6%	13.9%-15.3%	10.1%-20.4%
Race					
Black or African American	1,597	219	13.7%	12.1%-15.5%	10.3%-38.9%
White	14,094	2,000	14.2%	13.6%-14.8%	7.6%-19.4%
Other	3,911	450	11.5%	10.5%-12.6%	0.0%-23.5%
Unknown	134	10	7.5%	3.6%-13.3%	0.0%-12.3%
Ethnicity					
Hispanic or Latino	1,149	126	11.0%	9.2%-12.9%	0.0%-16.7%

Non-Hispanic	12,876	1,887	14.7%	14.1%-15.3%	10.9%-19.5%
Unknown / Unmapped	5,711	666	11.7%	10.8%-12.5%	0.0%-50.0%

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 (for example, hospitals). We examined confidence intervals around the estimates and variation in performance rates between **Hospitals 1-6** 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 **Hospitals 1-6** was 13.6% (95% CI: 13.1%, 14.1%). The performance rate ranged from 8.2% to 19.5% across all hospitals.

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 **Hospitals 1-6** showed performance scores that were within the range of harm rates found in the literature.[7] There was variation shown in the rate of harm across the six hospitals in these datasets, demonstrating a quality signal and suggesting room for improvement in rates of severe hyperglycemia among admitted patients.

References:

7. Maynard GA, Childers D, Holdych J, Kendall H, Hoag T, Harrison K. Improving Glycemic Control Safely in Non-Critical Care Patients: A Collaborative Systems Approach in Nine Hospitals. *Jt Comm J Qual Patient Saf.* 2017;43(4):179-188.

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 data for each required EHR data element for measure calculation.

For the EHR data elements used in this eCQM, 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 5 displays the data element reliability as the percent of missing data identified during adjudication for each data element required for measure calculation for **Hospitals 5-7**.

Table 5. Frequency of Missing Data by Data Element Required for Measure (Hospitals 5-7)

Data Element	Hospital 5 (N=175)		Hospital 6 (N=175)		Hospita
	Missing Count (#)	Missing Percent (%)	Missing Count (#)	Missing Percent (%)	Missing Count (#)
Admission date and time (mm/dd/yyyy, hh:mm)	0	0.0	0	0.0	0
Discharge date and time (mm/dd/yyyy, hh:mm)	0	0.0	0	0.0	0
Diabetes diagnosis	0	0.0	0	0.0	0
Medication administered, antidiabetic medication name	0	0.0	0	0.0	0
Laboratory test and point-of-care blood glucose results with date and time (mm/dd/yyy hh:mm, xx)	0	0.0	0	0.0	0

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*)

Among the data elements required for the measure calculation, the missing rate of all required data elements was 0%. This shows that it was feasible to extract the data elements for this eCQM from each hospital's EHR. Notably, while hospital 7 did have POC glucose lab data accurately available in structured fields and captured as part of workflow, it was not able to be appropriately mapped.

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).

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: Hospital_Harm_Hyperglycemia_Feasibility_Scorecard-637075211214470286.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.

N/A. This measure is not instrument-based.

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)
Public Reporting Payment Program 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 endorsement review and is not currently used in any accountability programs.

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

This eCQM is not currently publicly reported or used in any accountability programs. However, this measure is being developed for the Hospital Inpatient Quality Reporting (HIQR) and the Promoting Interoperability (PI) for Eligible Hospitals and Critical Access Hospitals programs pending NQF endorsement, MAP pre-rulemaking evaluation, and the CMS rulemaking process.

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.*)

N/A; the measure is not in current use. This measure is being developed for the Hospital Inpatient Quality Reporting (HIQR) and the Promoting Interoperability (PI) for Eligible Hospitals and Critical Access Hospitals programs pending NQF endorsement, MAP pre-rulemaking evaluation, and the CMS rulemaking process. CMS also may consider this measure for the Hospital-Acquired Conditions Reduction Program (HAC-RP) at some point in the future pending NQF endorsement, pre-rulemaking and rulemaking processes.

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

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

N/A; this measure is not publicly reported nor used in any accountability applications. Implementation is planned pending NQF endorsement, MAP pre-rulemaking evaluation, and the CMS rulemaking process.

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 not publicly reported nor used in any accountability applications. Implementation is planned pending finalization of the NQF endorsement 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 not publicly reported nor used in any accountability applications. Implementation is planned pending finalization of the NQF endorsement, MAP pre-rulemaking considerations, and CMS rulemaking processes.

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

N/A; this measure is not publicly reported nor used in any accountability applications. Implementation is planned pending finalization of the NQF endorsement 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 since it has not been implemented and is being submitted to NQF for endorsement, our team sought input from multiple stakeholder groups throughout the measure development cycle. We follow a transparent measure development process and highly value the feedback received on the measure. During this process, a technical expert panel composed of a variety of stakeholders was engaged at various stages of the development to obtain balanced, expert input. We also solicited and received feedback on the measure through an MMS Blueprint Public Comment 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 the TEP was instrumental in 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 newly developed eCQM, which is a significant departure from the retired predecessor measure on glycemic control, so there is no time trend information available regarding facility performance improvement. This eCQM is not currently in use in any quality improvement programs, but a primary goal of is measure 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 measure development or testing. However, we are committed to monitoring this measure's use and assessing potential unintended consequences for patients over time.

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

No unexpected benefits were noted during measure 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.

No

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

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

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):
Are the measure specifications harmonized to the extent possible?
5a.2. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.
N/A

5b. Competing Measures

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

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

Appendix

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

Attachment:

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): Centers for Medicare & 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: Stacie, Schilling, nqf@impaqint.com, 443-259-5133-

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

Kevin Kavanagh, MD, MS, Health Watch USA

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

Timothy Lowe, PhD, Premier Inc.

Christine (Chris) 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, FACP, RWJBarnabas Health and Rutgers University

J. Matthew Austin, MS, PhD, John Hopkins Medicine

Ann Borzecki, MD, MPH, Department of Veteran´s Affairs

John Bott, MSSW, MBA, The Leapfrog Group

Kyle Bruce, DPM, MPH, 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, BA, The Leapfrog Group

Richard Dutton, MD, MBA, United States Anesthesia Partners

Marybeth Foglia, RN, PhD, MA, Veterans Health Administration

Jeff Giullian, MD, MBA, DaVita Kidney Care

Maryellen Guinan, JD, America´s Essential Hospitals

Kate Kovich, MS, OTL, CPPS, Advocate Health Care

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

Karen Lynch, E, BSN, Massachusetts General Hospital

Milisa Manojlovich, PhD, RN, University of Michigan

Barbara Pelletreau, RN, MPH, Dignity Health

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

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

Brett Stauffer, MD, MHS, Baylor Scott and White Health

Brooks Udelsman, MD/MHS, Massachusetts General Hospital

Boback Ziaeian, MD, PhD, UCLA

Similar to the TEP, these Technical Advisory Group members responded to the posted Call for TEP members. The Technical Advisory Group was used in a manner 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? We anticipate annual updates and potentially triannual endorsement maintenance cycles.

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 user convenience. Users of proprietary code sets should obtain all necessary licenses from the owners of the 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 New Haven Health Service Corporation Center for Outcomes Research and Evaluation, and by 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.