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 #: 3503e

Measure Title: Hospital Harm – Severe Hypoglycemia

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

Brief Description of Measure: This electronic clinical quality measure (eCQM) assesses the proportion of inpatient admissions for patients aged 18 years and older who received at least one antihyperglycemic medication during their hospitalization, and who suffered a severe hypoglycemic event (blood glucose less than 40 mg/dL) within 24 hours of the administration of an antihyperglycemic agent.

Developer Rationale: This safety eCQM relates to glycemic control and hypoglycemia management in the hospital inpatient setting. Rates of inpatient hypoglycemic events are considered an indicator of the quality of care provided by a hospital. Hypoglycemic events are an adverse outcome that causes patients to experience drowsiness, confusion, anxiety, or irritability; sweating, weakness, increased heart rate, or trembling, as well as loss of consciousness, seizure or death.[1,2] Several important benefits related to quality improvement can be envisioned with the implementation of this eCQM. Furthermore, this eCQM will encourage providers to implement interventions aimed at better glycemic control and prevent severe hypoglycemia for hospital inpatients. In addition to avoiding direct patient harm from the severe hypoglycemic event, lower rates of hypoglycemia among hospitalized individuals would be expected to result in shorter lengths of stay and lower mortality.[3] Adoption of this performance eCQM has the potential to improve quality of care for individuals at risk of hypoglycemia and, therefore, advance the quality of care in the area of patient safety, 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 current hypoglycemia measure in a 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.[4]

Numerator Statement: The number of inpatient admissions during which a test for blood glucose with a result less than 40 mg/dL (severe hypoglycemia) where the event follows the administration of an antihyperglycemic medication within 24 hours.

Denominator Statement: All patients 18 years or older at the start of the encounter with a discharged inpatient hospital admission during the measurement period who were given at least one

antihyperglycemic medication during their hospital stay. The measure includes inpatient admissions which began in the Emergency Department or in observation status.

Denominator Exclusions: N/A, there are no denominator exclusions.

Measure Type: Outcome

Data Source: Electronic Health Records

Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date: N/A

Preliminary Analysis: New Measure

Criteria 1: Importance to Measure and Report

1a. Evidence

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

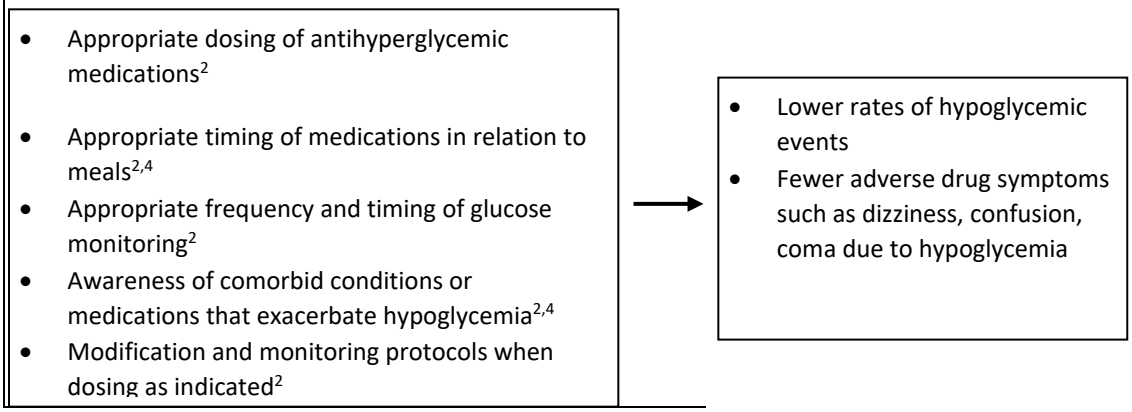
Evidence Summary or **Summary**

The goal of the Severe Hypoglycemia Electronic Clinical Quality Measure (eCQM) is to improve patient safety and prevent severe hypoglycemia in patients who are at risk. The focus of this outcome measure is inpatient hypoglycemia. The purpose of measuring hypoglycemic events is to reduce the frequency of these adverse patient outcomes and to improve hospitals' practices for appropriate dosing of medication and adequate monitoring of patients receiving glycemic control agents. Rates of inpatient hypoglycemic events can be reduced with high quality of care provided by a hospital.[1,2] Severe hypoglycemic events are largely avoidable by careful use of antihyperglycemic medication, monitoring of patient blood glucose levels, enhanced use of technology, and implementation of evidence-based best practices.[3,4,5,6,7,8]

Several important benefits related to quality improvement are envisioned with the implementation of this measure. Specifically, the measure will help providers identify individuals who develop severe hypoglycemia in the hospital inpatient setting. Furthermore, this eCQM will encourage providers to develop interventions to improve glycemic control for hospital inpatients, before a patient becomes hypoglycemic. In addition to avoiding direct patient harm from the hypoglycemic event, lower rates of hypoglycemia among hospitalized individuals would be expected to result in shorter lengths of stay and lower mortality. Moreover, the rate of severe hypoglycemia varies across hospitals indicating an opportunity for improvement in care. Hypoglycemic rates have been reported from 2.3% to 5% of hospitalized patients,[1,9] and from 0.4% of non-ICU patient days to 1.9% of ICU patient days.[3]

Hypoglycemic events are an adverse outcome that causes patients to experience a range of symptoms. The first signs of hypoglycemia include increased heart rate, sweating, uncontrollable trembling,

confusion, anxiety, and irritability. As blood glucose levels further decrease, the severity of symptoms increase, resulting in drowsiness, weakness, loss of consciousness, seizure, and coma.[8,10] Measuring this adverse event can improve hospitals' practices and reduce the occurrence of hypoglycemic events.



Updates:

Question for the Committee:

- *Is there at least one action 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 (Box 2) -> (yes) -> PASS

RATIONALE:

- Severe hypoglycemic events are largely avoidable by careful use of antihyperglycemic medications, monitoring of patient blood glucose levels, enhanced use of technology, and a hypoglycemia reduction bundle that includes creation of an internal hypoglycemia prevention task force to raise awareness of hypoglycemia risks and rates, with the development of data analytics.11,12,13 The Joint Commission and Johns Hopkins Hospital's Glucose Steering Committee outlined policies for inpatient glucose management, revolving around educating staff on the importance of glycemic management, disseminating best practices, and evaluating intervention effectiveness by using process and outcome measures.14 The American Diabetes Association (ADA) guidelines call for frequent blood glucose monitoring to properly implement diabetes therapy.15
- This Hospital harm – Severe Hypoglycemia eCQM provides a path to directly engage staff and hospital executives in the importance of glycemic measurement and will be a tool for quality improvement staff to assess internal metrics, along with providing CMS an instrument to assess the quality of care delivered to patients at risk for severe hypoglycemia across all acute care hospitals. Measuring hypoglycemic events in the hospital setting can help improve quality of care by identifying patients who develop hypoglycemia and incentivize hospitals to implement clinical workflows that facilitate evidence-based management to reduce the likelihood of severe hypoglycemic events.16 This eCQM has the potential to make care safer by reducing harm caused in the delivery of care, which is a National Quality Forum (NQF) healthcare priority.17

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

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

Maintenance measures – increased emphasis on gap and variation

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

This eCQM was tested with 2 test sites (6 hospitals) in 2 states (located in Midwest, South). Hospitals varied in size (50-1,000 beds, and 200-3,800 beds) and EHR system (Cerner, Epic), and were both teaching hospitals in urban settings. A detailed breakdown of the characteristics of the measured facilities and the patient population can be found in the attached Measure Testing Form (Beta Datasets 1 and 2).

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

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

- Number of Hospitals: 4
- Data collection period: Discharges between 1/1/2017 - 12/31/2017
- Denominator: 7,748
- Numerator: 195
- Performance rate: 2.52%
- 95% confidence interval: 2.18%, 2.89%
- Standard Deviation: 1.20%

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

- Number of Hospitals: 2
- Data collection period: Discharges between 1/1/2017 - 12/31/2017
- Denominator: 5,888
- Numerator: 174
- Performance rate: 2.96%
- 95% confidence interval: 2.54%, 3.42%
- Standard Deviation: 0.30%

Overall Performance

- Number of Hospitals: 6
- Performance Rate: 2.71%
- 95% confidence interval: 2.44%, 2.99%
- Standard deviation: 1.00%
- Range: 1.05% to 3.56%

Disparities

Data below are from initial development testing; this eCQM is not yet implemented.

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

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

- Number of Hospitals: 4
- Data collection period: 1/1/2017 - 12/31/2017
- Denominator (admissions): 7,748

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

- Number of Hospitals: 2
- Data collection period: 1/1/2017 - 12/31/2017
- Denominator: 5,888

Across Sites (n= 13,636, 6 hospitals)

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

18-64//7,529//206//2.7% (2.4%, 3.1%)

65+//6,107//163//2.7% (2.3%, 3.1%)

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

Male//7,130//197//2.8% (2.4%, 3.2%)

Female//6,487//170//2.6% (2.3%, 3.0%)

Unknown//19//2//10.5% (1.3%, 33.1%)

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

Black or African American//2,967//114//3.8% (3.2%, 4.6%)

White//8,386//188//2.2% (2.0%, 2.6%)

Other//2,011//55//2.7% (2.1%, 3.6%)

Unknown//272//12//4.4% (2.3%, 7.6%)

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

Hispanic or Latino//1,080//30//2.8% (1.9%, 3.9%)

Non-Hispanic//12,201//330//2.7% (2.4%, 3.0%)

Unknown//355//9//2.5% (1.2%, 4.8%)

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

Medicare//7,161//192//2.7% (2.3%, 3.1%)

Medicaid//1,359//46//3.4% (2.5%, 4.5%)

Private Insurance//4,225//110//2.6% (2.1%, 3.1%)

Self-pay or Uninsured//171//3//1.8% (0.4%, 5.0%)

Other (such as other government plans)//617//17//2.8% (1.6%, 4.4%)

Unknown//103//1//1.0% (0.0%, 5.3%)

Questions for the Committee:

- Is there a gap in care that warrants a national performance measure?
- Given the disparities data, should this measure be stratified or risk adjusted for SES?

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

RATIONALE:

- Range of performance across six hospitals was: 1.05% to 3.56%.

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

1a. Evidence
Comments:
**appropriate evidence
**Solid evidence for need and ability to impact frequency of event targeted; latter cleraly is harmful. While frequency is not high (< 3%), the harm and the wide potential population make the measure need evidence solid also

1b. Performance Gap
Comments:
**demonstrated performance gap
**Variation across tested site not high, but impact on health clear - so I see opportunity. There may be an race based disparity seen - that will be clearer with more data.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

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

Reliability

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

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

Validity

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

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

Composite measures only:

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

eCQM Technical Advisor(s) review:

Submitted measure is an HQMF compliant eCQM	The submitted eCQMspecifications follow the industry accepted format for eCQM (HL7 Health Quality Measures Format (HQMF)).
Documentation of HQMF,QDM, or CQL limitations	N/A – All components in the measure logic of the submitted eCQM are represented using the HQMF,QDM, or CQL standards
Value Sets	The submitted eCQM specifications uses existing value sets when possible and uses new value sets that have been vetted through the VSAC
Measure logic is unambiguous	Submission includes test results [from a simulated data set] demonstrating the measure logic can be interpreted precisely and unambiguously. – this includes 100% coverage of measured patient population testing with pass/fail test cases for each population
Feasibility Testing	Number of data elements included in measure calculation: 9 Number of data elements scoring less than 3 on scorecard: 0 All data elements assessed to be feasible across all domains (availability, accuracy, standards, and workflow)

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

Evaluators: NQF Scientific Methods Panel Subgroup

Methods Panel Review (Combined)

Evaluation of Reliability and Validity:

Scientific Methods Panel Votes: Measure passes

- Reliability: H-2, M-2, L-0, I-0
- Validity: H-1, M-3, L-0, I-0

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

Reliability

- Reliability was tested at the score level.
- *Note: Data element validity testing was also performed and is discussed in the validity section below. Per NQF guidance, data element validity testing is also acceptable for demonstrating data element reliability.*
- The developer calculated a signal to noise ratio (SNR) based on Adams' beta-binomial model.
    - The testing sample was 6 hospitals (13,636 eligible encounters) from Beta Datasets 1 and 2.
    - Median reliability score of 0.889 (range: 0.815-0.924)
- Reviewers agreed the score-level reliability results showed high agreement, but there was an initial concern that there is an insufficient number of hospitals to compute measure score reliability.
    - At least one panel member suggested that the sample size of facilities is appropriate since the beta-binomial reliability approach was used.

Validity

- Testing included empirical data element and score-level testing.
- Data Element
    - Data element testing evaluated the accuracy of electronically extracted EHR data elements compared with manually chart abstracted data elements from the same patients.
    - Two data sets were used for testing, which included six hospitals (two different EHR vendors). Beta Dataset 1 had 175 encounters, 97 being admissions with harm events and 78 being admissions without a harm event (denominator-only); and Beta Dataset 2 had 175 encounters, 100 being admissions with harm events and 75 being admissions without a harm event (denominator-only).
    - Data elements tested: Admission date and time; Antihyperglycemic medication administered; blood glucose test with date, time, result; and birth date.

- All but one data element (at one site) had a match rate over 95%, indicating valid and accurate data elements were extracted from the EHR. The exception was for antihyperglycemic medication (81%).
- One reviewed expressed concern that the "antihyperglycemic medication administered" element sensitivity was only 81%, but overall the Panel accepted the data element validity testing results.

- Score Level Empirical Testing
  - To validate the EHR-extracted numerator against the patient medical chart, to assess whether the harms actually occurred and captured the intended outcome, the developer clinically adjudicated each admission that met the criteria for a harm among the sample of abstracted records and calculated the positive predictive value (PPV) for all numerator and denominator cases.
    - Positive Predictive Value (PPV) of the numerator was 95%-99.2% and denominator was 98.9-100% across three data sets (n=11 hospitals).
  - Sensitivity, specificity, kappa, and negative predictive value (NPV) were calculated for Beta data sets 1 and 2 (n=6 hospitals).
    - Sensitivity is 100% in both data sets.
    - Specificity is 95% and 94%.
    - Kappa (95% CI) is 0.95 (CI 0.91,1.0) and 0.94 (CI 0.89, 1.0)
    - NPV is 100% in both data sets.
  - *Note: Per NQF criteria and the panel's discussion, the score level testing provided might be more appropriately considered additional data-element validity. One reviewer pointed out that there may be an argument that since the score is a sum of harm events, data element validity assures score level validity.*

- Face Validity
  - 10 out of 11 TEP members responded to "the measure as specified can be used to distinguish between better and worse quality care at hospitals" as follows: Moderately Disagreed (1), Somewhat Disagreed (1), Somewhat Agreed (0), Moderately Agreed (5), and Strongly Agreed (3).
  - *Note: typically NQF is less concerned with face validity assessments methods/results if empirical testing was conducted.*

- Reviewers' initial concerns included that the sample size was fairly limited (6 hospitals) and the lack of adjustment or stratification.
  - Regarding risk adjustment, the developer states that harms such as severe hypoglycemia should be avoidable regardless of patient risk. The developer notes that there is evidence indicating that most severe hypoglycemia events (<40mg/DL)

are avoidable, common causes are controllable in the hospital environment, and risk can be reduced using best practices.

- One panel member expressed concern that the measure is capturing a low-prevalence event. The developer responded that the measure is specifically focused on an at-risk population and noted the measure captures the critical outcome of severe hypoglycemic events.

- Meaningful differences
  - One reviewer shared that while there is some variation the sample is limited. Another reviewer suggested additional data be presented to determine meaningful differences.

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

Questions for the Committee regarding reliability:

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

Questions for the Committee regarding validity:

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

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

Methods Panel Evaluation (Combined): Scientific Acceptability

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 3503e

Measure Title: Hospital Harm – Severe Hypoglycemia

Type of measure:

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

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

Data Source:

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

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

☐ **Enrollment Data** ☐ **Other**

Level of Analysis:

☐ **Clinician: Group/Practice** ☐ **Clinician: Individual** ☒ **Facility** ☐ **Health Plan**

☐ **Population: Community, County or City** ☐ **Population: Regional and State**

☐ **Integrated Delivery System** ☐ **Other**

Measure is:

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

RELIABILITY: SPECIFICATIONS

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

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

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

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

   Methods Panel member 1:Measure description (numerator, denominator, and exclusions) seem appropriate. Why would a patient would experience a severe hypoglycemic event while in the hospital after receiving an antihyperglycemic medication?

   Methods Panel member 2: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 <u>VALIDITY</u> testing** of <u>patient-level data</u> conducted?

   ☐ **Yes** ☐ **No** Methods Panel member 1: (X) NA—score level reliability conducted

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

   Submission document: Testing attachment, section 2a2.2

   Methods Panel member 2:Adams' beta-binomial approach used appropriately.

   Methods Panel member 1:Very few hospitals (n=11) were included in the measure testing. The total number of patients seem sufficient, but the measure is at the facility level. Score level reliability = beta-binomial method of signal-to-noise (ratio of variances between providers).

   Methods Panel member 3:Calculated a signal-to-noise ratio.
   Methods Panel member 4:SNR based on the beta-binomial model

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

Submission document: Testing attachment, section 2a2.3

Methods Panel member 2:Based on 13,636 eligible encounters across 6 hospitals in Beta Datasets 1 and 2, the signal-to-noise ratio yielded a median reliability score of 0.889 (range: 0.815-0.924), which indicates excellent agreement.

Methods Panel member 1: Measure score reliability seems to be measured at the data element level. "There were 13,636 eligible encounters across 6 hospitals in Beta Datasets 1 and 2. The signal-to-noise ratio yielded a median reliability score of 0.889 (range: 0.815-0.924)." Measure score is reported at the hospital level. Only 6 hospitals were used to compute reliability? Which 6 of the 11?

Methods Panel member 3:Calculated a signal-to-noise ratio of 0.89. This indicates excellent agreement.

Methods Panel member 4:Median reliability was excellent (0.89). However, in the absence of risk adjustment, it is unknown to what extent the variation between providers represents true variation in quality versus variation simply due to differences in patient case mix between providers.

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** Methods Panel member 1: Insufficient # of hospitals

☒ **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.

Methods Panel member 2:Appropriate score level testing was conducted with strong results.

Methods Panel member 1: Insufficient number of hospitals to compute measure score reliability

Methods Panel member 3:Score-level testing was done; used appropriate method; found high level of agreement.
Methods Panel member 4:In the absence of risk adjustment, it is unknown to what extent the variation between providers represents true variation in quality versus variation due to differences in patient case mix between providers. It is not enough for MD to simply indicate that this complication can be avoided using best practices. The MD needs to provide empiric evidence that patient characteristics, such as age, diabetes, frailty, history of stroke are not associated with the development of pressure ulcers. The high level of score-level reliability may simply reflect the lack of risk adjustment. This is, of course, an empiric question. The MD needs to demonstrate that risk adjustment is not necessary.
Methods Panel member 5:Signal to noise, high

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

Submission document: Testing attachment, section 2b2.

Methods Panel member 2:None.

Methods Panel member 1: No measure exclusions.

Methods Panel member 3:Not applicable.

Methods Panel member 4:none

Methods Panel member 5: N/A

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

Submission document: Testing attachment, section 2b4.

Methods Panel member 2:None.

Methods Panel member 1: No data presented that show meaningful differences or how these would be computed by the developer.

Methods Panel member 3:See some variation, but have a fairly limited sample size (n-=6).

Methods Panel member 4:The lack of risk adjustment makes it impossible to determine if "measured" differences in performance reflect true differences in quality.

Methods Panel member 5:No

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

Submission document: Testing attachment, section 2b5.

Methods Panel member 2:None.
Methods Panel member 1: Different electronic data systems have different requirements for level of detail that may jeopardize data element reliability/quality.
Methods Panel member 3:Not applicable.
Methods Panel member 5: None.

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

Submission document: Testing attachment, section 2b6.

Methods Panel member 2:None.

Methods Panel member 1: Developers believe that there may be some missing data. However, given that the measure is not risk adjusted or stratified in reporting, there would be minimal impact of missing data.

Methods Panel member 3:No concerns.

Methods Panel member 5: None

16. Risk Adjustment

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

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

☐ Yes ☒ No ☐ Not applicable Methods Panel member 3:Indicated that performance is modifiable if a hospital follows best practice.

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 Methods Panel member 1: (X) NA—measure is not risk adjusted

16d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☐ Yes ☐ No Methods Panel member 1: (X) NA—measure is not risk adjusted

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

Methods Panel member 1: (X) NA—measure is not risk adjusted

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

☐ Yes ☐ No

Methods Panel member 1: (X) NA—measure is not risk adjusted

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

Methods Panel member 1: (X) NA—measure is not risk adjusted

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

Methods Panel member 1: I disagree with the decision to not risk adjust and/or the decision to not stratify the reported results. There are clear differences among patient populations served among hospitals, including percentages of patients with diabetes, and certain racial groups that are more likely to have diabetes. To compare quality performance across hospitals, some consideration of risk adjustment or stratification should be made.

Methods Panel member 4: There is no risk adjustment.

VALIDITY: TESTING

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

18. Method of establishing validity of the measure score:

☒ **Face validity**

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

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

19. Assess the method(s) for establishing validity

Submission document: Testing attachment, section 2b2.2

Methods Panel member 2:Data element validity testing wsa performed in two "Beta datasets" drawn from 6 hospitals. From these hospitals, a stratified random sample of total admissions were selected, including 97 and 175 patients respectively. Trained abstrators extracted all of the case information from EMRs at each site and these were compared to the data used to calculate the emeasure.

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

Methods Panel member 1: Narrative describing data element validity methodology was confusing (e.g., discussion of "simulating a series of *moe* and target PPV values"). Table 3 presentation of results of methodology provided clearer information.

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

Methods Panel member 4:The agreement between EMR and chart was tested by re-abstraction.

Methods Panel member 3:For score-level validity, demonstrated both empirical validty testing and face validty. The empirical validity testing, they compared EHR extraction and manual chart abstraction. For face validity, asked a TEP about the measure bing a good measure of hospital quality. For data-element validty, they compared the sensitivity of the data elements (EHR structured fields vs. manual abstraction).

20. Assess the results(s) for establishing validity

Submission document: Testing attachment, section 2b2.3

Methods Panel member 2:All but one data element (at one site) had a match rate over 95%, indicating valid and accurate data elements were extracted from the EHR. The exception was for antihyperglycemic medication (81%) administered in Beta Dataset 1, which the developers explain as a documentation issue.

All three datasets had a PPV over 95%, meaning in almost all the cases the admission met the criteria for a harm in both the chart abstracted and EHR-extracted data.

Methods Panel member 1:

Table 3 seems to show reasonable results for data element validity, except for date of birth.

Table 4 seem to show reasonable results. However, number of hospitals for which these results apply is missing and is necessary to see if the strong results are generalizable.

Methods Panel member 3:Score-level validty testing was excellent for empirical testing (PPV and NPV >95%) and face validity (80% of TEP was in agreement that measure was good measure of quality).

Some concerns with data-element validity testing and the Antihyperglycemic medication administered sensitivity only being 81%.

Methods Panel member 4:Assessed the validity of outcome data element using sensitivity, specificity, NPV, and kappa statistic. These measures of agreement were excellent.

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

Submission document: Testing attachment, section 2b1.

☒ **Yes** Methods Panel member 1: if the number of hospitals is sufficient

☐ **No**

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

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

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

☒ **Yes** Methods Panel member 1: if the number of hospitals is sufficient

☐ **No**

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

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

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

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

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

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

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

Methods Panel member 2: Appropriate measures were used with good results.

Methods Panel member 1: Concerns about number of hospitals used to compute measure score require a lower limit score.

Methods Panel member 4:The lack of risk adjustment is a critical limitation of this measure. A priori, it would be unreasonable to assume that patient frailty and comorbidities do not play an important role in the incidence of severe hypoglycemia. A frail patient with caloric malnutrition is more likely to develop hypoglycemia than a middle-aged patient with type II DM. Similarly, there is a spectrum of diabetic patients, some of whom are much more brittle than others. Finally, postoperative patients experience a stress response which also may make it more difficult to achieve adequate glucose control, and avoid hypoglycemia. Ultimately, the need for risk adjustment is an empiric question that needs to be explored by the MD. It is not sufficient to simply "affirm" that no risk adjustment is necessary because this complication can be prevented using best practices.

Methods Panel member 3:Have some concerns with the data-element validty testing and the finding that the sensitivity of the Antihyperglycemic Med was 81%.

Methods Panel member 5: Data element: compare EHR with charts, PPV
Score: sensitivity, specificity, kappa and NPV

ADDITIONAL RECOMMENDATIONS

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

Methods Panel member 1: There are common problems across all of the "hospital harm" electronic measures that need to be addressed.

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

2a1. Reliability – Specifications
Comments:
**no concerns
**Testing is strong and supports reliability of accessing and categorizing key information in the variable.

2a2. Reliability – Testing
Comments:
**None
**No. Looks to be moderate to high in assessing the current information.

2b1. Validity –Testing
Comments:
**None
**Same - no concerns

2b4-7. Threats to Validity
2b4. Meaningful Differences
Comments:
**None
**Given the construct, it is possible episodes could be missed - but that is likely a rare event. I have no systematic validity threats given the 40 mg/dL threshold definition of the event.

2b2-3. Other Threats to Validity
2b2. Exclusions
2b3. Risk Adjustment
Comments:
**none

**No risk adjusting called for or needed; social factors are not primary to the measure though may impact effects of events.

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent

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

- Data are generated from electronic health records in the regular course of care.
- The eCQM team did not identify any feasibility issues with this measure.

Questions for the Committee:

- Are the required data elements routinely generated and used during care delivery?
- Are the required data elements available in electronic form, e.g., EHR or other electronic sources?
- Is the data collection strategy ready to be put into operational use?
- If an eCQM, does the eCQM Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites?

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

RATIONALE:

- Data are generated from electronic health records in the regular course of care.

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

3. Feasibility
Comments:
**no concerns
**No concerns - these data are readily avvailble, shown in beta testing.

Criterion 4: Usability and Use

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

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

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

4a.1. Accountability and Transparency. Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after

initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

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

OR

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

Accountability program details

- This eCQM is not currently publicly reported or used in an accountability application because it has only recently completed re-specification and is being submitted to NQF for endorsement in its re-specified form. The previously NQF-endorsed measure was not implementable because the MAT could not support the measure as specified when it was originally developed. The measure was re-specified using the updates to the MAT including expression of the logic with CQL. This re-specified measure was presented to the Measure Applications Partnership (MAP) in December 2018 and received conditional support for rulemaking, pending NQF review and endorsement.

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

- While this measure does not have usability information from measured entities, as it has been re-specified as an eCQM and has not been implemented yet, our team sought input from multiple stakeholder groups throughout the measure development process. We believe in a transparent measure development process and highly value the feedback received on the measure. During the development, 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 44-day Public Input Period during development.

Additional Feedback: N/A

***Questions for the Committee**:*

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

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

RATIONALE:

- Planned for use in accountabilty program per the measure developer.

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 [Impact/trends over time/improvement]

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 None reported by developer

Potential harms None reported by 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:

- No identified unintended consequences.

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

4a1. Use - Accountability and Transparency

Comments:
**no concerns
**Clear plans for sharing and feedback loops exist and will be key if adopted and later continued.

4b1. Usability – Improvement

Comments:
**none
**Little real harm assessment done - that is, excessive hyperglycemia. This will need to be maintenance focus, and current benefit assessment is clear and adequate.

Criterion 5: Related and Competing Measures

Related or competing measures

- No competing measures identified by the developer.

Harmonization

- N/A

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

5. Related and Competing
Comments:
**None

Public and Member Comments

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

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

Developer Submission

1 Brief Measure Information

NQF #: 3503e

Corresponding Measures:

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

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

De.3. Brief Description of Measure: This electronic clinical quality measure (eCQM) assesses the proportion of inpatient admissions for patients aged 18 years and older who received at least one antihyperglycemic medication during their hospitalization, and who suffered a severe hypoglycemic event (blood glucose less than 40 mg/dL) within 24 hours of the administration of an antihyperglycemic agent.

1b.1. Developer Rationale: This safety eCQM relates to glycemic control and hypoglycemia management in the hospital inpatient setting. Rates of inpatient hypoglycemic events are considered an indicator of the quality of care provided by a hospital. Hypoglycemic events are an adverse outcome that causes patients to experience drowsiness, confusion, anxiety, or irritability; sweating, weakness, increased heart rate, or trembling, as well as loss of consciousness, seizure or death.[1,2] Several important benefits related to quality improvement can be envisioned with the implementation of this eCQM. Furthermore, this eCQM will encourage providers to implement interventions aimed at better glycemic control and prevent severe hypoglycemia for hospital inpatients. In addition to avoiding direct patient harm from the severe hypoglycemic event, lower rates of hypoglycemia among hospitalized individuals would be expected to result in shorter lengths of stay and lower mortality.[3] Adoption of this performance eCQM has the potential to improve quality of care for individuals at risk of hypoglycemia and, therefore, advance the quality of care in the area of patient safety, 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 current hypoglycemia measure in a 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.[4]

References

1. Classen, D. C., Jaser, L., & Budnitz, D. S. (2010). Adverse drug events among hospitalized Medicare patients: Epidemiology and national estimates from a new approach to surveillance. Jt Comm J Qual Patient Saf, 36(1), 12-21.

2. American Diabetes Association. Hypoglycemia (Low Blood Glucose). 2015; http://diabetes.org/living-with-diabetes/treatment-and-care/blood-glucose-control/hypoglycemia-low-blood.html. Accessed August 20, 2018.

3. Nirantharakumar K, Marshall T, Kennedy A, Narendran P, Hemming K, Coleman JJ. Hypoglycaemia is associated with increased length of stay and mortality in people with diabetes who are hospitalized. Diabet Med. 2012;29(12):e445-448.

4. Services USDoHaH. National Action Plan for Adverse Drug Event Prevention. Washington, DC2014.

S.4. Numerator Statement: The number of inpatient admissions during which a test for blood glucose with a result less than 40 mg/dL (severe hypoglycemia) where the event follows the administration of an antihyperglycemic medication within 24 hours.

S.6. Denominator Statement: All patients 18 years or older at the start of the encounter with a discharged inpatient hospital admission during the measurement period who were given at least one antihyperglycemic medication during their hospital stay. The measure includes inpatient admissions which began in the Emergency Department or in observation status.

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? N/A

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

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

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

Hospital_Harm_Severe_Hypoglycemia_NQF_Evidence_Submission_Form.docx

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

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

Yes

Evidence (subcriterion 1a)

[NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*): Previously endorsed as 2363e, now submitted as 3503e

Measure Title: Hospital Harm – Severe Hypoglycemia

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Not applicable
Date of Submission: 4/2/2019

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 Hypoglycemia

☐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 enter measure title

☐ Process: Click here to name what is being measured

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

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

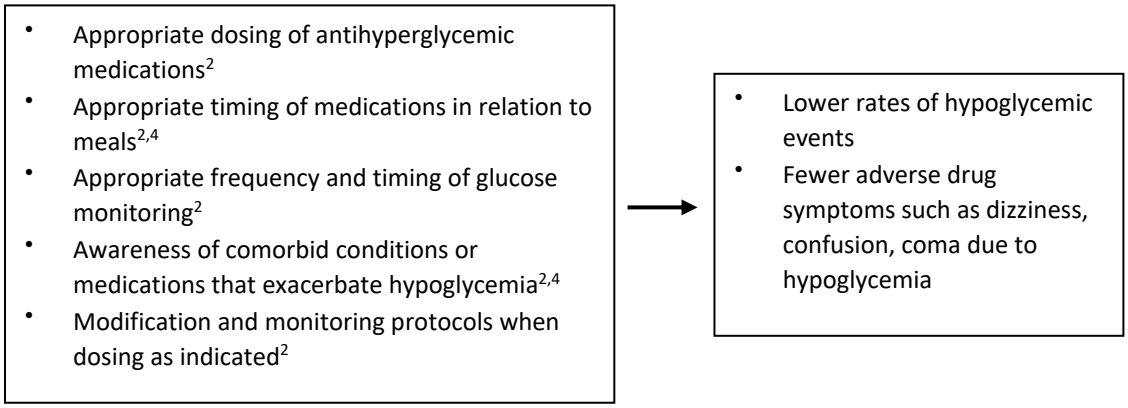
1a.2 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process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

The goal of the Severe Hypoglycemia Electronic Clinical Quality Measure (eCQM) is to improve patient safety and prevent severe hypoglycemia in patients who are at risk. The focus of this outcome measure is inpatient hypoglycemia. The purpose of measuring hypoglycemic events is to reduce the frequency of these adverse patient outcomes and to improve hospitals' practices for appropriate dosing of medication and adequate monitoring of patients receiving glycemic control agents. Rates of inpatient hypoglycemic events can be reduced with high quality of care provided by a hospital.[1,2] Severe hypoglycemic events are largely avoidable by careful use of antihyperglycemic medication, monitoring of patient blood glucose levels, enhanced use of technology, and implementation of evidence-based best practices.[3,4,5,6,7,8]

Several important benefits related to quality improvement are envisioned with the implementation of this measure. Specifically, the measure will help providers identify individuals who develop severe hypoglycemia in the hospital inpatient setting. Furthermore, this eCQM will encourage providers to develop interventions to improve glycemic control for hospital inpatients, before a patient becomes hypoglycemic. In addition to avoiding direct patient harm from the hypoglycemic event, lower rates of hypoglycemia among hospitalized individuals would be expected to result in shorter lengths of stay and lower mortality. Moreover, the rate of severe hypoglycemia varies across hospitals indicating an opportunity for improvement in care. Hypoglycemic rates have been reported from 2.3% to 5% of hospitalized patients,[1,9] and from 0.4% of non-ICU patient days to 1.9% of ICU patient days.[3]

Hypoglycemic events are an adverse outcome that causes patients to experience a range of symptoms. The first signs of hypoglycemia include increased heart rate, sweating, uncontrollable trembling, confusion, anxiety, and irritability. As blood glucose levels further decrease, the severity of symptoms increase, resulting in drowsiness, weakness, loss of consciousness, seizure,

and coma.[8,10] Measuring this adverse event can improve hospitals' practices and reduce the occurrence of hypoglycemic events.



References:

1. Wexler DJ, Meigs JB, Cagliero E, Nathan DM, Grant RW. Prevalence of hyper- and hypoglycemia among inpatients with diabetes: a national survey of 44 U.S. hospitals. *Diabetes Care.* 2007;30(2):367-369.
2. American Diabetes Association. 14. Diabetes Care in the Hospital: Standards of Medical Care in Diabetes—2018. *Diabetes Care.* 2018;41(Supplement 1):S144.
3. 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.
4. Moghissi ES, Korytkowski MT, DiNardo M, et al. American Association of Clinical Endocrinologists and American Diabetes Association Consensus Statement on Inpatient Glycemic Control. *Diabetes Care.* 2009;32(6):1119-1131.
5. Office of the Inspector General (OIG), US Department of Health and Human Services. *Adverse Events in Hospitals: National Incidence Among Medicare Beneficiaries.* 2010.
6. American Diabetes Association. Hypoglycemia (Low Blood Glucose). 2015; http://diabetes.org/living-with-diabetes/treatment-and-care/blood-glucose-control/hypoglycemia-low-blood.html. Accessed August 20, 2018.
7. Milligan PE, Bocox MC, Pratt E, Hoehner CM, Krettek JE, Dunagan WC. Multifaceted approach to reducing occurrence of severe hypoglycemia in a large healthcare system. *Am J Health Syst Pharm.* 2015;72(19):1631-1641.
8. Jeffrey Schnipper CL, Chima Ndumele, and Merri Pendergrass. Effects of a Computerized Order Set on the Inpatient Management of Hyperglycemia: A Cluster-Randomized Controlled Trial. *Endocrine Practice.* 2010;16(2):209-218.
9. Nirantharakumar K, Marshall T, Kennedy A, Narendran P, Hemming K, Coleman JJ. Hypoglycaemia is associated with increased length of stay and mortality in people with diabetes who are hospitalized. *Diabet Med.* 2012;29(12):e445-448.
10. Classen DC, Jaser L, Budnitz DS. Adverse drug events among hospitalized Medicare patients: epidemiology and national estimates from a new approach to surveillance. *Jt Comm J Qual Patient Saf.* 2010;36(1):12-21.

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

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

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

Severe hypoglycemic events are largely avoidable by careful use of antihyperglycemic medications, monitoring of patient blood glucose levels, enhanced use of technology, and a hypoglycemia reduction bundle that includes creation of an internal hypoglycemia prevention task force to raise awareness of hypoglycemia risks and rates, with the development of data analytics.[11,12,13] The Joint Commission and Johns Hopkins Hospital's Glucose Steering Committee outlined policies for inpatient glucose management, revolving around educating staff on the importance of glycemic management, disseminating best practices, and evaluating intervention effectiveness by using process and outcome measures.[14] The American Diabetes Association (ADA) guidelines call for frequent blood glucose monitoring to properly implement diabetes therapy.[15]

This Hospital harm – Severe Hypoglycemia eCQM provides a path to directly engage staff and hospital executives in the importance of glycemic measurement and will be a tool for quality improvement staff to assess internal metrics, along with providing CMS an instrument to assess the quality of care delivered to patients at risk for severe hypoglycemia across all acute care hospitals. Measuring hypoglycemic events in the hospital setting can help improve quality of care by identifying patients who develop hypoglycemia and incentivize hospitals to implement clinical workflows that facilitate evidence-based management to reduce the likelihood of severe hypoglycemic events.[16] This eCQM has the potential to make care safer by reducing harm caused in the delivery of care, which is a National Quality Forum (NQF) healthcare priority.[17]

References:

11. Milligan PE, Bocox MC, Pratt E, Hoehner CM, Krettek JE, Dunagan WC. Multifaceted approach to reducing occurrence of severe hypoglycemia in a large healthcare system. *Am J Health Syst Pharm.* 2015;72(19):1631-1641.
12. Jeffrey Schnipper CL, Chima Ndumele, and Merri Pendergrass. Effects of a Computerized Order Set on the Inpatient Management of Hyperglycemia: A Cluster-Randomized Controlled Trial. *Endocrine Practice.* 2010;16(2):209-218.
13. Greg Maynard KK, Pedro Ramos, Diana Childers, Brian Clay, Meghan Sebasky, Ed Fink, Aaron Field, Marian Renvall, Patricia S. Juang, Charles Choe, Diane Pearson, Brittany Serences, and Suzanne Lohnes. Impact of a Hypoglycemia Reduction Bundle and a Systems Approach to Inpatient Glycemic Management. *Endocrine Practice.* 2015;21(4):355-367.
14. Munoz M, Pronovost P, Dintzis J, et al. Implementing and Evaluating a Multicomponent Inpatient Diabetes Management Program: Putting Research into Practice. *The Joint Commission Journal on Quality and Patient Safety.* 2012;38(5):195-AP194.

15. American Diabetes Association. Standards of Medical Care in Diabetes. Diabetes Care. 2018; Supp 1: S1-S15. http://care.diabetesjournals.org/content/41/Supplement_1.
16. Aloi JA, Mulla C, Ullal J, Lieb DC. Improvement in Inpatient Glycemic Care: Pathways to Quality. *Current Diabetes Reports.* 2015;15(4):18.
17. National Quality Forum. *Prioritization of High-Impact Medicare Conditions and Measure Gaps: Measure Prioritization Advisory Committee Report.* Washington, DC: NQF;2010.

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?	

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

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

1a.4 OTHER SOURCE OF EVIDENCE

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

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

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

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

1b. Performance Gap

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

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

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

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

This safety eCQM relates to glycemic control and hypoglycemia management in the hospital inpatient setting. Rates of inpatient hypoglycemic events are considered an indicator of the quality of care provided by a hospital. Hypoglycemic events are an adverse outcome that causes patients to experience drowsiness, confusion, anxiety, or irritability; sweating, weakness, increased heart rate, or trembling, as well as loss of consciousness, seizure or death.[1,2] Several important benefits related to quality improvement can be envisioned with the implementation of this eCQM. Furthermore, this eCQM will encourage providers to implement interventions aimed at better glycemic control and prevent severe

hypoglycemia for hospital inpatients. In addition to avoiding direct patient harm from the severe hypoglycemic event, lower rates of hypoglycemia among hospitalized individuals would be expected to result in shorter lengths of stay and lower mortality.[3] Adoption of this performance eCQM has the potential to improve quality of care for individuals at risk of hypoglycemia and, therefore, advance the quality of care in the area of patient safety, 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 current hypoglycemia measure in a 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.[4]

References

1. Classen, D. C., Jaser, L., & Budnitz, D. S. (2010). Adverse drug events among hospitalized Medicare patients: Epidemiology and national estimates from a new approach to surveillance. Jt Comm J Qual Patient Saf, 36(1), 12-21.

2. American Diabetes Association. Hypoglycemia (Low Blood Glucose). 2015; http://diabetes.org/living-with-diabetes/treatment-and-care/blood-glucose-control/hypoglycemia-low-blood.html. Accessed August 20, 2018.

3. Nirantharakumar K, Marshall T, Kennedy A, Narendran P, Hemming K, Coleman JJ. Hypoglycaemia is associated with increased length of stay and mortality in people with diabetes who are hospitalized. Diabet Med. 2012;29(12):e445-448.

4. Services USDoHaH. National Action Plan for Adverse Drug Event Prevention. Washington, DC2014.

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

This eCQM was tested with 2 test sites (6 hospitals) in 2 states (located in Midwest, South). Hospitals varied in size (50-1,000 beds, and 200-3,800 beds) and EHR system (Cerner, Epic), and were both teaching hospitals in urban settings. A detailed breakdown of the characteristics of the measured facilities and the patient population can be found in the attached Measure Testing Form (Beta Datasets 1 and 2).

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

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

- Number of Hospitals: 4
- Data collection period: Discharges between 1/1/2017 - 12/31/2017
- Denominator: 7,748
- Numerator: 195
- Performance rate: 2.52%
- 95% confidence interval: 2.18%, 2.89%
- Standard Deviation: 1.20%

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

- Number of Hospitals: 2
- Data collection period: Discharges between 1/1/2017 - 12/31/2017
- Denominator: 5,888
- Numerator: 174
- Performance rate: 2.96%
- 95% confidence interval: 2.54%, 3.42%
- Standard Deviation: 0.30%

Overall Performance

- Number of Hospitals: 6
- Performance Rate: 2.71%
- 95% confidence interval: 2.44%, 2.99%
- Standard deviation: 1.00%
- Range: 1.05% to 3.56%

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

N/A

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for maintenance of endorsement. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

Data below are from initial development testing; this eCQM is not yet implemented.

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

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

- Number of Hospitals: 4
- Data collection period: 1/1/2017 - 12/31/2017
- Denominator (admissions): 7,748

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

- Number of Hospitals: 2
- Data collection period: 1/1/2017 - 12/31/2017
- Denominator: 5,888

Across Sites (n= 13,636, 6 hospitals)

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

18-64//7,529//206//2.7% (2.4%, 3.1%)

65+//6,107//163//2.7% (2.3%, 3.1%)

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

Male//7,130//197//2.8% (2.4%, 3.2%)

Female//6,487//170//2.6% (2.3%, 3.0%)

Unknown//19//2//10.5% (1.3%, 33.1%)

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

Black or African American//2,967//114//3.8% (3.2%, 4.6%)

White//8,386//188//2.2% (2.0%, 2.6%)

Other//2,011//55//2.7% (2.1%, 3.6%)

Unknown//272//12//4.4% (2.3%, 7.6%)

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

Hispanic or Latino//1,080//30//2.8% (1.9%, 3.9%)

Non-Hispanic//12,201//330//2.7% (2.4%, 3.0%)

Unknown//355//9//2.5% (1.2%, 4.8%)

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

Medicare//7,161//192//2.7% (2.3%, 3.1%)

Medicaid//1,359//46//3.4% (2.5%, 4.5%)

Private Insurance//4,225//110//2.6% (2.1%, 3.1%)

Self-pay or Uninsured//171//3//1.8% (0.4%, 5.0%)

Other (such as other government plans)//617//17//2.8% (1.6%, 4.4%)

Unknown//103//1//1.0% (0.0%, 5.3%)

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

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

N/A

2. Reliability and Validity—Scientific Acceptability of Measure Properties

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

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

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

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

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

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

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

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

This is an eMeasure **Attachment:** Del18c2HOP5HarmsHypoITS12172018v5_6_Artifacts-636824656414337046.zip,Hypoglycemia_Bonnie_Test_Cases_Results.pdf

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

Attachment **Attachment:** Del18c2HOP5HarmsHypoFeasibilityScorecard12172018_v02.xlsx

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

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

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

Not an instrument-based measure

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

Yes

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

This measure is a re-specification of a previously NQF-endorsed measure that has never been used in a CMS program. Changes to the measure specifications are as follows:

Numerator differences: The current Hospital Harm—Severe Hypoglycemia measure assesses whether a severe hypoglycemia event occurred during an inpatient hospitalization (dichotomous outcome). The previous NQF-endorsed measure counted number of hypoglycemia events in the numerator per patient days in the denominator.

Additionally, the Hospital Harm–Severe Hypoglycemia measure assesses the use of specific antihyperglycemic medications found in the Value Set Authority Center (VSAC) that are likely to cause hypoglycemia, within 24 hours of administration. The measure no longer has separate specifications for short-acting insulin. The previous NQF-endorsed measure differentiated between administration of short-acting insulin within 12-hours and other medications within 24 hours.

These changes will ease the burden on hospitals and be meaningful to patients, while still adhering to the original intent of the measure.

Denominator differences: The current Hospital Harm–Severe Hypoglycemia measure examines the total number of admissions with at least one antihyperglycemic agent administered during the hospital stay. The NQF-endorsed measure examined the total number of hospital days with at least one anti-diabetic agent administered.

This change aligns with the numerator change to number of admissions, which eases hospital burden.

Exclusions differences: The current Hospital Harm–Severe Hypoglycemia measure specifications do not have any denominator exclusions; the previous NQF-endorsed measure excludes admissions with lengths of stay greater than 120 days.

This exclusion was dropped as it is not applicable to the current measure specifications because the measure is not based on patient days.

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

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

The number of inpatient admissions during which a test for blood glucose with a result less than 40 mg/dL (severe hypoglycemia) where the event follows the administration of an antihyperglycemic medication within 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 data to calculate the measure score. The time period for data collection is during an inpatient hospitalization, beginning at hospital arrival (whether through Emergency Department, observation stay, or directly admitted as inpatient).

All data elements necessary to calculate this measure are defined within value sets available in the 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 both laboratory and point-of-care glucose tests, including venous or arterial blood and serum or plasma.

The antihyperglycemic medications are defined by the value set of Hypoglycemics (2.16.840.1.113762.1.4.1179.3). This value set includes medications and insulin capable of causing hypoglycemia in a patient.

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

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

All patients 18 years or older at the start of the encounter with a discharged inpatient hospital admission during the measurement period who were given at least one antihyperglycemic medication during their hospital stay. The measure includes inpatient admissions which began in the Emergency Department or in observation status.

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

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

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

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

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

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

Encounters who were given at least one antihyperglycemic medication are defined by the value set of Hypoglycemics (2.16.840.1.113762.1.4.1179.3), which also defines the numerator medications. This value set includes medications and insulin capable of causing hypoglycemia in a patient.

To access the value sets for the measure, 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 measure is not stratified.

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

No risk adjustment or risk stratification

If other:

S.12. Type of score:

Rate/proportion

If other:

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

Better quality = Lower score

S.14. Calculation Algorithm/Measure Logic (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)*

Target population: Inpatient admission encounters, all payer, where individuals are aged 18 years or older at the start of the admission and who were given at least one antihyperglycemic medication during their hospital stay, within the measurement period.

To create the denominator:

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

2. Determine the patient's age in years. The patient's age is equal to the admission date minus the birth date. If the patient is 18 years or older, go to Step 3. If less than 18 years old, do not include in the measure population.

3. Determine if there was at least one antihyperglycemic medication (from the Hypoglycemic value set 2.16.840.1.113762.1.4.1179.3) administered during the inpatient hospitalization (including in the Emergency Department or observation stay if later converted into an inpatient admission). If not, do not include in the measure population.

To create the numerator, for each encounter identify:

1. Any instance of a test for blood glucose with a result less than 40 mg/dL during the encounter is considered a severe hypoglycemic event, including values from either laboratory or Point of Care (POC) testing.

2. For any value less than 40mg/dL, determine if there was an antihyperglycemic medication administered by hospital staff within the 24 hours before the event and during the hospitalization (including emergency department and observation stays contiguous with the admission). If not, do not include in the numerator.

a. The 24-hour time frame extends from the end of the medication administration to the start of the blood glucose test.

3. For any value less than 40mg/dL, do not include any events (identified in Step 1) if it was followed by a repeat POC test for blood glucose within 5 minutes of the initial test and with a result greater than 80 mg/dL.

a. Rationale: The measure logic does –not– require a repeat blood glucose test to be performed. The expectation is that in most cases of severe hypoglycemia, the clinical team will be treating the patient and will not immediately repeat the test. However, if the severe hypoglycemic event is suspected to be spurious, for example if the patient is clinically asymptomatic, and a repeat test is performed to confirm that suspicion, this step will remove false positives that can occur in POC testing to ensure hospitals are not penalized for erroneous results. The 5-minute time frame extends from the time that the initial blood glucose test was performed to the time that the repeat blood glucose test was performed.

Only the first qualifying severe hypoglycemic event is counted in the numerator, and only one severe hypoglycemic event is counted per encounter.

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

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

N/A; this measure does not use a sample 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 measure 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 MAT output, which includes the human readable and XML artifacts of the clinical quality language (CQL) for the measure are contained in the eCQM specifications attached. No additional tools are used for data collection for eCQMs.

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

No data collection instrument provided

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

Facility

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

Inpatient/Hospital

If other:

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

N/A

2. Validity – See attached Measure Testing Submission Form

Del18c2HOP5HarmsHypoITSForm010219-636824679941320611.docx,Del18c2HOP5HarmsHypoTestingForm012219_v0.2.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.

Yes

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.

Yes

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.

No - This measure is not risk-adjusted

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

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

Measure Number (*if previously endorsed*)**:** 2363
Measure Title: Hospital Harm – Severe Hypoglycemia
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 specificaitons** (e.g., claims and EHRs), section **2b5** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b1-2b6) must be in this

form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.

- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 25 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for version 7.1 of the Measure Testing Attachment.

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

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

2b1. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **instrument-based measures (including PRO-PMs) and composite performance measures**, validity should be demonstrated for the computed performance score.

2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure; [12]

AND

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

2b3. For outcome measures and other measures when indicated (e.g., resource use):

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

OR

- rationale/data support no risk adjustment/ stratification.

2b4. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;

OR

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

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

2b6. Analyses identify the extent and distribution of **missing data** (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

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

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

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

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

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

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

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

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

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

Measure Specified to Use Data From: (*must be consistent with data sources entered in S.17*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ claims	☐ claims

☐ registry	☐ registry
☐ abstracted from electronic health record	☒ abstracted from electronic health record
☒ eMeasure (HQMF) implemented in EHRs	☒ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

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

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

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

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

1.4. What levels of analysis were tested? (*testing must be provided for <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 datasets, dates, number of measured hospitals, and number of admissions used in each phase of testing are in **Table 1**.

Table 1. Dataset Descriptions

Dataset	Applicable Section in the Testing Attachment	Description of Dataset	EHR Vendor
Beta Dataset 1	Section 2a2 Reliability Testing Section 2b1 Validity Testing Section 2b4 Identification of Statistically Significant and Meaningful Differences in Performance Section 2b6 Missing Data Analysis	Dates of Data: January 1, 2017 - December 31, 2017 Number of Hospitals: 4 Number of Admissions: 7,748 Number of Unique Patients: 5,394 For Validity Testing: sample of 175 admissions Hospitals were within one health system, in urban locations. Hospitals ranged between 50 – 1,000 beds. Located in the Midwest.	Epic
Beta Dataset 2	Section 2a2 Reliability Testing Section 2b1 Validity Testing Section 2b4 Identification of Statistically Significant and Meaningful Differences in Performance	Dates of Data: January 1, 2017 - December 31, 2017 Number of Hospitals: 2 Number of Admissions: 5,888 Number of Unique Patients: 4,580 For Validity Testing: sample of 175 admissions Hospitals were within one health system, in urban	Cerner

	Section 2b6 Missing Data Analysis	locations. Hospitals ranged between 200 – 3,800 beds. Located in the South.	
Alpha Dataset	Section 2b1 Validity Testing (Measure Score)	Dates of Data: June 1, 2016 - May 31, 2017 Number of Hospitals: 5 Number of Admissions: 66,127 Hospitals were in two different health systems, both in urban locations, and not-for-profit. They were diverse in terms of bed size (between 100-199 beds and 300-399 bed), teaching status, geographic location (South, West).	Cerner & Epic

Patient descriptive characteristics in Alpha Dataset are as follows:

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

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

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

Initial Patient Population Characteristics	Beta Dataset 1 (N, %)	Beta Dataset 2 (N, %)	Across Beta Sites (N, %)
Number of unique patients	5394	4580	9974
Average Age [Mean(SD)]	59 (15)	65 (16)	62 (15)
18-35	415, 7.7%	244, 5.3%	659, 6.6%
36-64	2929, 54.3%	1827, 39.9%	4756, 47.7%
65+	2050, 38.0%	2509, 54.8%	4559, 45.7%

Initial Patient Population Characteristics	Beta Dataset 1 (N, %)	Beta Dataset 2 (N, %)	Across Beta Sites (N, %)
Sex			
Male	2930, 54.3%	2304, 50.3%	5234, 52.5%
Female	2464, 45.7%	2263, 49.4%	4727, 47.4%
Unknown	0, 0.0%	13, 0.3%	13, 0.1%
Race			
Black or African-American	1377, 25.5%	638, 13.9%	2015, 20.2%
White	3729, 69.1%	2458, 53.7%	6187, 62.0%
Other	288, 5.3%	1269, 27.7%	1557, 15.6%
Unknown	0, 0.0%	215, 4.7%	215, 2.2%
Ethnicity			
Hispanic or Latino	82, 1.5%	781, 17.1%	863, 8.7%
Non-Hispanic	5279, 97.9%	3582, 78.2%	8861, 88.8%
Unknown	33, 0.6%	217, 4.7%	250, 2.5%
(Primary) Payer			
Medicare	2826, 52.4%	2276, 49.7%	5102, 51.2%
Medicaid	572, 10.6%	412, 9.0%	984, 9.9%
Private Insurance	1732, 32.1%	1451, 31.7%	3183, 31.9%
Self-pay or Uninsured	4, 0.1%	131, 2.9%	135, 1.4%
Other (such as other government plans)	178, 3.3%	310, 6.8%	488, 4.9%
Unknown	82, 1.5%	0, 0.0%	82, 0.8%

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

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

As described in Section 1.7, Table 2, we collected information on the following social risk factors using data extracted from hospital EHR systems: race, ethnicity, and primary payer.

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)
☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)
☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

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

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

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

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

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

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

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

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

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

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

Yu, H, Mehrota, A, Adams J. (2013). Reliability of utilization measures for primary care physician profiling. Healthcare, 1, 22-29.
Adams, J. The Reliability of Provider Profiling: A Tutorial. Santa Monica, CA: RAND Corporation, 2009. https://www.rand.org/pubs/technical_reports/TR653.html.

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

Measure Score Reliability Results
There were 13,636 eligible encounters across 6 hospitals in **Beta Datasets 1 and 2**. The signal-to-noise ratio yielded a median reliability score of 0.889 (range: 0.815-0.924).

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 0.89 indicates excellent agreement.

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

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

2b1. VALIDITY TESTING

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

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

☒ **Performance measure score**

☒ **Empirical validity testing**

☒ **Systematic assessment of face validity of 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 the purpose of these analyses.

Data Element Validity:

For **Beta Datasets 1 and 2**, a stratified sample of 175 total admissions were selected at each hospital test site. Sample size calculations ensure a robust sample was used for validity testing. Specifically, we derived our sample size based on the following assumptions: Our primary endpoint for sample size estimation is PPV, which is applicable for both data element validity and measure score validity. We adjudicated all our numerator cases in alpha test and obtained high PPVs (>90% in most of the cases). Based on this, we approximate the sample size based on one-sample proportion formula as the following:

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

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

Beta Dataset 1 had 175 encounters, 97 being admissions with harm events and 78 being admissions without a harm event (denominator-only); and **Beta Dataset 2** had 175 encounters, 100 being admissions with harm events and 75 being admissions without a harm event (denominator-only). Data were abstracted from the EHR by trained abstractors at each test site; abstractors at all sites had experience abstracting data for chart-based quality measure reporting. Abstractors were provided with an instruction manual and an Access database to document the information abstracted from the EHR. Access databases were only pre-populated with the unique patient identifier; abstractors were asked to input all other data from the chart independently of the EHR dataset. Abstraction training was also provided to each site.

Table 3 shows the 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 **Beta Datasets 1 and 2**. Each data element matched if the specific electronically extracted value exactly matched the manually abstracted value (gold standard). For example, out of 84 specific instances where a patient was administered a antihyperglycemic medication (in the chart data), 68 of those specific cases were extracted correctly in the EHR data, resulting in an 81% match rate. For date/time data elements, we matched month, day, year, hour, and minutes. For glucose lab values, we matched on the glucose value result (whole integers), date, time +/- one minute. For administration of an antihyperglycemic medications, we matched on the name of the medication ordered, as the timestamps of medication administered in the EHR were autogenerated, and not found as easily in the chart.

Empirical Measure Score Validity

Measure score validity assesses whether the harm rate (or, the measure score outcome) calculated for each facility is in fact accurate. The measure score is calculated for each facility

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

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

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

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

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 hypoglycemic events obtained from the Hospital Harm – Severe Hypoglycemia Measure 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 3. Data Element Validity (Sensitivity) Results Required for Measure (Beta Datasets 1 and 2)

Data Element	Beta Dataset 1			Beta Dataset 2		
	# 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%
Antihyperglycemic medication administered: order ID	68	84	81.0%	40	41	97.6%
Laboratory test, blood glucose test with date, time, result (mm/dd/yyy hh:mm XXX)	2118	2215	95.6%	2341	2454*	95.4%
Patient characteristic: birth date (mm/dd/yyyy)	175	168	96.0%**	175	175	100.0%**

*Data element validity for glucose result in **Beta Dataset 2** were matched using point-of-care tests only, due to human error by abstractor.
**Patient date of birth was assessed using PPV percent match, as # cases matched in abstraction / # cases in EHR.

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

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

Measure	Beta Dataset 1				Beta Dataset 2			
	Sensitivity	Specificity	Kapa (95% CI)	NPV	Sensitivity	Specificity	Kappa (95% CI)	NPV
Severe Hypoglycemia	100%	95%	0.95 (0.91, 1.0)	100%	100%	94%	0.94 (0.89, 1.0)	100%

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

Measure Component	Alpha Dataset PPV	Beta Dataset 1 PPV	Beta Dataset 2 PPV
Initial patient population/ Denominator	N/A	100.0%	98.9%
Numerator	99.2%	95.9%	95.0%

Face Validity

10 out of 11 TEP members responded to the survey question as follows: Moderately Disagreed (1), Somewhat Disagreed (1), Somewhat Agreed (0), Moderately Agreed (5), and Strongly Agreed (3). The two TEP members who disagreed somewhat and moderately did not provide rationale.

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

Data Element Validity

All but one data element (at one site) had a match rate over 95%, indicating valid and accurate data elements were extracted from the EHR. The exception was for antihyperglycemic medication (81%) administered in **Beta Dataset 1**. We believe this specific match rate was due to different naming conventions between the way medications names are stored in the EHR and a basic drop-down list used by the chart abstracted data.

For the blood glucose date, time, and result data element, we assessed the validity of all glucoses recorded during the hospitalization, for a more robust sample to evaluate a clearer picture of data element accuracy. The match rate is below 100% due to the analysis matching on all three fields (date, time and result), which is a higher standard than matching of individual elements, and the timing of each variables does not match exactly within one minute. However, match rates were still high at 95.6% and 95.4% respectively. Overall, we believe the data elements required for the measure show validity.

Empirical Measure Score Validity
All three datasets had a PPV over 95%, meaning that almost all cases, the admission 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 both Beta Dataset 1 and 2 and specificity is 95% and 94% in Beta Dataset 1 and 2 respectively. This means that the probability of the EHR data detecting a true hypoglycemic event in patients that had a true hypoglycemic event based on the abstracted data ('gold standard') is 100% (sensitivity). The probability of the EHR data detecting no hypoglycemia out of the no hypoglycemic event based on abstracted data is 94-95% (specificity). NPV was 100% in both Beta Dataset 1 and 2, indicating the EHR data indicated a harm did not occur, and 100% of the time the chart abstraction confirmed a harm did not occur. Kappa of 0.94 and 0.95 indicate excellent agreement. We will continue to reevaluate validity through reevaluation as hospitals participate in this measure and as required by NQF for maintenance of endorsement.

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

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

Face Validity:
80% of TEP members agreed (moderately or strongly) that the measure will provide an accurate reflection of quality, which reflects good face validity.

2b2. EXCLUSIONS ANALYSIS
NA ☒ no exclusions — *skip to section 2b3*

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

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 degree of risk, it may be important to adjust for patient risk factors in order to compare hospital performance. However, many harms such as severe hypoglycemia should be avoidable, regardless of patient risk. We consider the following criteria in determining whether risk adjustment is warranted for the severe hypoglycemia measure:

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 patients' inherent risk profile. For example, there may be evidence that even complex patients with multiple risk factors can avoid harm events when providers closely adhere to care guidelines

In the case of the severe hypoglycemia eCQM, there is evidence indicating that most hypoglycemic events of this severity (<40 mg/DL) are avoidable. Although certain patients may be particularly vulnerable to hypoglycemia in certain settings (e.g. due to organ failure and not related to administration of diabetic agents), the most common causes are lack of caloric intake, overuse of anti-diabetic agents, or both. 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 measure. 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 measure, we examined the performance (harm) rate of the measure across patient characteristics of age, sex, race, ethnicity, and payer. Age (by date of birth) was validated; no other patient demographic was validated using chart data. It is important to note these results are derived from a small dataset that is not generalizable to the entire population and the datasets include many characteristics that are 'unknown' in the EHR which limits the usability of the results; additionally, we do not believe it is clinically appropriate to adjust by these characteristics given the clinical rationale provided above.

Table 6. Performance Rate by Encounter Characteristic (Beta Dataset 1 and 2)

Characteristic	Beta Dataset 1			Beta Dataset 2			Across Beta Sites		
	Denominator	Numerator	Performance Rate % (95% CI)	Denominator	Numerator	Performance Rate % (95% CI)	Denominator	Numerator	Performance Rate % (95% CI)
Number of unique Encounters	7,748	195	2.5 (2.2, 2.9)	5,888	174	3.0 (2.5, 3.4)	13,636	369	2.71 (2.4, 3.0)
Average Age									
18-64	4,882	119	2.4 (2.0, 2.9)	2,647	87	3.3 (2.6, 4.0)	7,529	206	2.7 (2.4, 3.1)
65+	2,866	76	2.7 (2.1, 3.3)	3,241	87	2.7 (2.2, 3.3)	6,107	163	2.7 (2.3, 3.1)
Sex									
Male	4,202	112	2.7 (2.2, 3.2)	2,928	85	2.9 (2.3, 3.6)	7,130	197	2.8 (2.4, 3.2)
Female	3,546	83	2.3 (1.9, 2.9)	2,941	87	3.0 (2.4, 3.6)	6,487	170	2.6 (2.3, 3.0)
Unknown	0	0	N/A	19	2	10.6 (1.3, 33.1)	19	2	10.5 (1.3, 33.1)
Race									
Black or African-American	2,158	80	3.7 (3.0, 4.6)	809	34	4.2 (2.9, 5.8)	2,967	114	3.8 (3.2, 4.6)
White	5,193	104	2.0 (1.6, 2.4)	3,193	84	2.6 (2.1, 3.3)	8,386	188	2.2 (2.0, 2.6)
Other	397	11	2.8 (1.4, 4.9)	1,614	44	2.7 (2.0, 3.6)	2,011	55	2.7 (2.1, 3.6)

Characteristic	Beta Dataset 1			Beta Dataset 2			Across Beta Sites		
	Denominator	Numerator	Performance Rate % (95% CI)	Denominator	Numerator	Performance Rate % (95% CI)	Denominator	Numerator	Performance Rate % (95% CI)
Unknown	0	0	N/A	272	12	4.4 (2.3, 7.6)	272	12	4.4 (2.3, 7.6)
Ethnicity									
Hispanic or Latino	112	2	1.8 (0.2, 6.3)	968	28	2.9 (1.9, 4.2)	1,080	30	2.8 (1.9, 3.9)
Non-Hispanic	7,599	193	2.5 (2.2, 2.9)	4,602	137	3.0 (2.5, 3.5)	12,201	330	2.7 (2.4, 3.0)
Unknown	37	0	0.0 (0.0, 0.0)	318	9	2.8 (1.3, 5.3)	355	9	2.5 (1.2, 4.8)
(Primary) Payer									
Medicare	4,145	100	2.4 (2.0, 2.9)	3,016	92	3.1 (2.5, 3.7)	7,161	192	2.7 (2.3, 3.1)
Medicaid	792	23	2.9 (1.9, 4.3)	567	23	4.1 (2.6, 6.0)	1,359	46	3.4 (2.5, 4.5)
Private Insurance	2,468	68	2.8 (2.2, 3.5)	1,757	42	2.4 (1.7, 3.2)	4,225	110	2.6 (2.1, 3.1)
Self-pay or Uninsured	5	0	0.0 (0.0, 52.2)	166	3	1.8 (0.4, 5.2)	171	3	1.8 (0.4, 5.0)
Other (such as other government plans)	235	3	1.3 (0.26, 3.7)	382	14	3.7 (2.0, 6.1)	617	17	2.8 (1.6, 4.4)
Unknown	103	1	1.0 (0.02, 5.3)	0	0	N/A	103	1	1.0 (0.02, 5.3)

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

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

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

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

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

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

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

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

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

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

2b3.9. Results of Risk Stratification Analysis:

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

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

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

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

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

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

The performance rate across all hospitals in both **Beta Datasets 1 and 2** was 2.71% (95% CI: 2.44%, 2.99%). The performance rate ranged from 1.05% to 3.56% across all hospitals in both datasets.

The performance rate for all hospitals in **Beta Dataset 1** was 2.52% (95% CI: 2.18%, 2.89%).

The performance rate for all hospitals in **Beta Dataset 2** was 2.96% (95% CI: 2.54%, 3.42%).

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

Results from **Beta Datasets 1 and 2** showed performance scores that were within range of harm rates found in the literature (Nirantharakumar and Marshall, 2012; Wexler and Meigs, 2007). There was variation shown in the rate of harm across the six hospitals in this dataset, demonstrating a quality signal, suggesting room for improvement in rates of severe hypoglycemia among admitted patients.

References:

Nirantharakumar, K., Marshall, T., Kennedy, A., Narendran, P., Hemming, K., & Coleman, J. J. (2012). Hypoglycaemia is associated
with increased length of stay and mortality in people with diabetes who are hospitalized. Diabetic Medicine, 29(12), e445-e448.
Wexler, D. J., Meigs, J. B., Cagliero, E., Nathan, D. M., & Grant, R. W. (2007). Prevalence of hyper- and hypoglycemia among
inpatients with diabetes: A national survey of 44 U.S. hospitals. Diabetes Care, 30(2), 367-369.

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

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

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

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

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

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

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

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

We quantitatively assessed data element feasibility using the rate of missing for each required EHR data element for measure calculation.

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

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

Table 7. Frequency of Missing Data by Data Element Required for Measure (Beta Datasets 1 and 2)

Data Element	Beta Dataset 1			Beta Dat	
	Missing Count (#)	Encounters (#)	Missing Percent (%)	Missing Count (#)	Encou
Admission characteristic: admission date and time	0	7,748	0.0%	0	5,88
Antihyperglycemic medication administered: order ID	0	7,748	0.0%	0	5,88
Antihyperglycemic medication administered with date and time	0	7,748	0.0%	0	5,88
Laboratory test, blood glucose results	0	7,748	0.0%	0	5,88
Laboratory test, blood glucose date and time	0	7,748	0.0%	0	5,88
Patient characteristic: birth date	0	7,748	0.0%	0	5,88

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, there were no missing data meaning all encounters had all required data elements, showing that it was feasible to extract the data elements from each test site's EHR. This means each encounter had an antihyperglycemic medication, and a blood glucose result. Because all patients in the measure denominator received at least one antihyperglycemic medication during their hospitalization, we expect all patients to receive glucose blood tests.

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

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

If other:

3b. Electronic Sources

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

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

ALL data elements are in defined fields in electronic health records (EHRs)

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

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

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

Attachment: Del18c2HOP5HarmsHypoFeasibilityScorecard12172018_v02-636892935277842876.xlsx

3c. Data Collection Strategy

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

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

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

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

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

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

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

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

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. Current and Planned Use

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

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

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

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

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

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

This eCQM is not currently publicly reported or used in an accountability application because it has only recently completed re-specification and is being submitted to NQF for endorsement in its re-specified form. The previously NQF-endorsed measure was not implementable because the MAT could not support the measure as specified when it was originally developed. The measure was re-specified using the updates to the MAT including expression of the logic with CQL. This re-specified measure was presented to the Measure Applications Partnership (MAP) in December 2018 and received conditional support for rulemaking, pending NQF review and endorsement.

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

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

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

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

N/A; this measure has been re-specified as an eCQM and as such has not been implemented. Implementation is planned pending finalization of the NQF and CMS rulemaking processes.

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

N/A; this measure has been re-specified as an eCQM and as such has not been implemented. Implementation is planned pending finalization of the NQF and CMS rulemaking processes.

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

Describe how feedback was obtained.

N/A; this measure has been re-specified as an eCQM and as such has not been implemented. Implementation is planned pending finalization of the NQF and CMS rulemaking processes.

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

N/A; this measure has been re-specified as an eCQM and as such has not been implemented. Implementation is planned pending finalization of the NQF and CMS rulemaking processes.

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

While this measure does not have usability information from measured entities, as it has been re-specified as an eCQM and has not been implemented yet, our team sought input from multiple stakeholder groups throughout the measure development process. We believe in a transparent measure development process and highly value the feedback received on the measure. During the development, 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 44-day Public Input Period during development.

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

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

Improvement

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

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

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

This is a re-specified eCQM and there is no time trend information available regarding facility performance improvement. This eCQM is not currently used in any quality improvement programs, but a primary goal of the 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 eCQM development or testing. However, we are committed to monitoring this measure's use and assessing its potential unintended consequences over time, such as the inappropriate shifting of care and other negative unintended consequences for patients.

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

No unexpected benefits were noted during eCQM development testing.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

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.

No appendix **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: Benjamin, Shirley, bshirley@impaqint.com, 202-774-1964-

Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

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

Technical Expert Panel Members:

David Baker, MD, MPH, The Joint Commission

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

Lisa Freeman, BA, Connecticut Center for Patient Safety

Patrick Guffey, MD, University of Colorado Department of Anesthesiology

David Hopkins, MS, PhD, Stanford University

Kevin Kavanagh, MD, MS, Health Watch USA

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

Timothy Lowe, PhD, Premier Inc.

Jennifer Meddings, MD, MSc, University of Michigan Health System

Christine Norton, MA, Patient/Consumer/Caregiver

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

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

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

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

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

Technical Advisory Group Members:

Andy Anderson, MD, MBA, RWJBarnabas Health and Rutgers University

Matt Austin, MS, PhD, John Hopkins Medicine

Ann Borzecki, MD, Department of Veteran´s Affairs

John Bott, The Leapfrog Group

Kyle Bruce, DPM, Riverbend Medical Group

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

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

Melissa Danforth, The Leapfrog Group

Richard Dutton, MD, Baylor University

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

Jeff Giullian, MD, MBA, DaVita Kidney Care

Maryellen Guinan, America´s Essential Hospitals

Kate Kovich, Advocate Health Care

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

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

Milisa Manojlovich, MD, University of Michigan

Barbara Pelletreau, Dignity Health

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

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

Brett Stauffer MD MHS FHM, Baylor Scott and White Health

Brooks Udelsman, MD/MHS, Massachusetts General Hospital

Boback Ziaeian, UCLA

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

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released:

Ad.3 Month and Year of most recent revision:

Ad.4 What is your frequency for review/update of this measure? We anticipate annual updates and potentially triennial 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.