NATIONAL QUALITY FORUM
Driving measurable health improvements together

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 #: 3687e

Corresponding Measures:

Measure Title: ePC-07 Severe Obstetric Complications

Measure Steward: The Joint Commission

Brief Description of Measure: Hospital-level measure scores are calculated as a risk-adjusted proportion of the number of delivery hospitalizations for women who experience a severe obstetric complication, as defined by the numerator, by the total number of delivery hospitalizations in the denominator during the measurement period. The hospital-level measure score will be reported as a rate per 10,000 delivery hospitalizations.

ePC07 was developed in collaboration with Yale New Haven Health Services Corporation/Center for Outcomes Research and Evaluation (CORE).

Developer Rationale: The United States experiences higher rates of maternal morbidity and mortality than most other developed countries. These rates have continued to trend upward in recent decades.1 Research indicates that the overall rate of severe maternal morbidity (SMM) has increased by almost 200% between 1993 and 2014 to 144 per 10,000 delivery hospitalizations1, with more than 25,000 women per year experiencing obstetric complications.2 Recent maternal mortality data from 2018 reveal that 658 women died from maternal causes, resulting in a rate of 17.4 deaths per 100,000 live births, with 77% of the deaths attributed to direct obstetric causes like hemorrhage, preeclampsia, obstetric embolism, and other complications.3 This has prompted national health experts and organizations to prioritize quality improvement strategies to mitigate risk of adverse outcomes among maternal populations. The U.S. Department of Health & Human Services (HHS) has also called for action to improve maternal health and outcomes and outlines seven actions for healthcare professionals, including participating in quality improvement and safety initiatives.4 There are currently only a small number of quality measures focused on maternal health, and those implemented at the national level are mostly process measures and limited in scope. While these existing measures aim to promote coordination of care and standardize health care processes, maternal health outcome measures are sorely needed. Measures that are focused on maternal health outcomes will address the patient safety priority area under the Meaningful Measures 2.0 framework, and likewise will use EHR data to address interoperability, another meaningful measure area for assessing quality of health care.5

1. Centers for Disease Control and Prevention. Severe Maternal Morbidity in the United States. January 31, 2020; https://www.cdc.gov/reproductivehealth/maternalinfanthealth/severematernalmorbidity.html.

Numerator Statement: Inpatient hospitalizations for patients with severe obstetric complications including the following:

- Severe maternal morbidity diagnoses (see list below)
- Severe maternal morbidity procedures (see list below)
- Discharge disposition = expired

Severe Maternal Morbidity Diagnoses:

- Cardiac
    - Acute heart failure
    - Acute myocardial infarction
    - Aortic aneurysm
    - Cardiac arrest/ventricular fibrillation
    - Heart failure/arrest during procedure or surgery
- Hemorrhage
    - Disseminated intravascular coagulation
    - Shock
- Renal
    - Acute renal failure
- Respiratory
    - Adult respiratory distress syndrome
    - Pulmonary edema
- Sepsis
- Other OB
    - Air and thrombotic embolism
    - Amniotic fluid embolism
    - Eclampsia
    - Severe anesthesia complications
- Other Medical
    - Puerperal cerebrovascular disorder
    - Sickle cell disease with crisis

Severe Maternal Morbidity Procedures:

- Blood transfusion
- Conversion of cardiac rhythm
- Hysterectomy
- Temporary tracheostomy
- Ventilation

For further details on changes made to the numerator specifications during pilot testing, please see Changes Made to ePC07 Specifications During Pilot Testing in additional attachments.

Denominator Statement: Initial Patient Population: Inpatient hospitalizations for patients age >= 8 years and < 65 admitted to the hospital for inpatient acute care who undergo a delivery procedure with a discharge date that ends during the measurement period

Denominator: Inpatient hospitalizations for patients delivering stillborn or live birth with >= 20 weeks, 0 days gestation completed

Denominator Exclusions: Patients with confirmed diagnosis of COVID with COVID-related respiratory condition or patients with confirmed diagnosis of COVID with COVID-related respiratory procedure.

For further details on changes made to the denominator exclusion specifications during pilot testing please see Changes Made to ePC07 Specifications During Pilot Testing in additional attachments.

Measure Type: Outcome

Data Source: Electronic Health Data; Electronic Health Records

Level of Analysis: Facility

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.

The developer provides the following description for this measure:

- This is a new eCQM measure at the facility level that calculates a risk adjusted rate of deliveries with a severe obstetrical complication.
- The developer provides a logic model that depicts hospital assessment of delivering persons for factors associated with maternal morbidity and mortality which leads to monitoring the rate of severe maternal complications/mortality. These two actions result in hospitals reviewing severe obstetric complication cases and incorporating quality improvement practices which ultimately leads to the reduction in severe obstetric outcomes and improved quality of life for obstetric patients and babies.

Summary:

- The developer presents empirical data from a journal articles and Maternal Mortality Review Committees to show the following:
  - Data suggests that a large portion of maternal mortality can be avoided. A 2019 report from 14 maternal mortality review committees determined that 65.8 percent of obstetric maternal deaths were preventable. Another study found that 40.5 percent of pregnancy-related deaths were preventable.
  - Data suggest much of severe maternal morbidity is similarly avoidable. A study found that 45.5 percent of near-miss morbidity and 16.7 percent of other severe morbidities were preventable.
- Areas that the provider can impact for prevention of pregnancy-related morbidity/mortality include: assessment/point of entry to care, diagnosis and recognition of high risk, referral to experts, treatment, management hierarchy, education, communication, policies and procedures, documentation and discharge (Geller et al, 2004).

Question for the Committee:

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

Guidance from the Evidence Algorithm

- Measure assesses performance on a health outcome -> Yes, Developer provides a relationship between the measured outcome and at least one healthcare action -> Yes -> Rate as PASS

Preliminary rating for evidence: ☒ Pass ☐ No Pass

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

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

- Data for 25 hospitals were used for 2020 discharges using a rate per 10,000 deliveries, and includes both mortality and morbidity.
    - The mean risk adjusted severe obstetric complications rate was 248.8 (standard deviation (SD) of 55.5). The other reported rates were as follows:
        - Low: 157.1
        - 25th percentile: 215.6
        - 75th percentile: 287.3
        - High: 369.5
- The developer supports this data with data from the literature showing that the United States experiences higher rates of severe obstetric complications than most other developed countries. The overall rate of severe maternal morbidity (SMM) has increased by almost 200 percent between 1993 and 2014 to 144 per 10,000 delivery hospitalizations. The U.S. Department of Health and Human Services has called for action to improve maternal health outcomes, including participation in quality improvement and safety initiatives.

Disparities

- The developer presents a study that states women who identify as racial and ethnic minority groups are at a significantly higher risk for developing severe obstetric complications than non-Hispanic White women.
- Using their testing data, the developer found that when adjusting for risk factors, Non-Hispanic African-American women have a significantly increased risk (18 percent) of having any SMM compared to non-Hispanic White women, while Hispanic women had a significantly increased risk (41 percent) and Non-Hispanic Asian/Pacific Islander women had a significantly increased risk (62 percent) for any SMM.
- When excluding blood transfusion-only cases, compared to non-Hispanic White women, non-Hispanic African-American women had a 6 percent increased risk of SMM, while Hispanic women had a 36 percent increased risk and non-Hispanic Asian/Pacific Islander women had a 43 percent increased risk.
- When compared to private insurance, Medicaid and Medicare covered beneficiaries also showed an increased risk when adjusting for risk factors for any SMM and SMM excluding blood transfusion-only cases.

Questions for the Committee:

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

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

Committee Pre-evaluation Comments:

1a. Evidence

- Yes, many cases of severe maternal morbidity are preventable through patient safety bundles and improved quality of care. However, certain cases are more preventable than others so the measure may need to be refined and condensed to a more restricted set of preventable indicators to be used in comparing hospital performance.
- Significant problem with opportunities for improved prevention
- New electronic measure. Good evidence to support measure focus.
- Evidence applies directly. For example, QBL is a process measure that leads to earlier recognition of an OB hemorrhage, possible prevention of an unplanned hysterectomy or even death.

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

- The disparities data presented is not consistent with national data showing a much larger Black-White gap https://www.hcup-us.ahrq.gov/faststats/SMMServlet?radio-2=on&location1=US&characteristic1=01C13&location2=&characteristic2=01C11&expansionInfoState=hide&dataTablesState=hide&definitionsState=hide&exportState=hide
- see variation in performance on the measure; noted racial disparities
- Significant performance gaps are documented.
- Variability exists across the nation in performance. Variation and disparities when comparing race and ethnicity exists when analyzing performance and complications in white persons compared to other race groups.

Criteria 2: Scientific Acceptability of Measure Properties

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

Evaluators: Christie Teigland, Alex Sox-Harris, Jack Needleman, Sean O'Brien, Jeff Geppert, Larry Glance, Marybeth Farquhar, Sherrie Kaplan, Terri Warholak, Sam Simon, Paul Kurlansky (Combined Methods Panel Review)

- The SMP passed this measure on Reliability with a score of: H-4; M-5; L-1; I-0
- The SMP passed this measure on Validity with a score of: H-2; M-6; L-0; I-2

2a. Reliability: Specifications and Testing

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

- Submitted measure specification follows established technical specifications for eCQMs (health quality measure format (HQMF), Quality Data Model (QDM), Clinical Quality Language (CQL)) as indicated Sub-criterion 2a1.
- Submitted measure specification is fully represented and is not hindered by any limitations in the established technical specifications for eCQMs.

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.

Specifications:

- Measure specifications are complex, but clear and precise.
- eCQMs was specified using the latest industry accepted eCQM technical specifications: QDM, HQMF, and CQL, and value sets vetted through the National Library of Medicine's Value Set Authority Center (VSAC).

Reliability Testing:

- Reliability testing was conducted at the Patient/Encounter Level:
  - The developer used patient/encounter validity testing to serve as patient/encounter reliability testing (please see validity section below).
- Reliability testing was conducted at the Accountable Entity Level:
  - Reliability testing was conducted with data from 8 pilot sites representing 25 individual hospitals who all had at least 25 deliveries per year, over the time period 1/1/20-12/31/20. Results were also calculated for hospitals with at least 200 deliveries per year (23 of the 25 hospitals). The developer evaluated accountable entity reliability using a signal-to-noise ratio.
  - At the health site level, median reliability was 0.991 (range of 0.982-0.997) for any severe obstetric complications and 0.955 (0.916-0.983) for severe obstetric complications excluding blood transfusion-only cases.
  - For hospitals with at least 25 delivery encounters, the median reliability score was 0.959 (0.802-0.996) for any severe obstetric complication outcome and 0.684 (0.273-0.961) for severe obstetric complications excluding blood transfusion-only cases.
  - The median reliability is higher when included hospitals had at least 200 delivery encounters for the year; the median reliability score was 0.978 (0.867-0.996) for any severe obstetric complication outcome and 0.804 (0.377-0.961) for severe obstetric complications excluding blood transfusion-only cases.

SMP Summary:

- Reliability testing was seen as acceptable, though it was noted that while the signal-to-noise reliability results indicate very high reliability, these results appear to change when blood transfusion cases are excluded.

Questions for the Committee regarding reliability:

- *Do you have any concerns that the measure cannot 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?*

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

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

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.

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

Validity Testing

- Validity testing conducted at the Patient/Encounter Level:
  - Validity testing was completed for 15 individual hospitals at 6 pilot sites. This includes one system of 10 hospitals and 5 individual hospitals. The developer reviewed 3-4 charts for each hospital in the system and 30-36 charts at the individual hospitals. The review included three different EHR vendors, 3 sites use Epic, 2 use Meditech and 1 site uses Cerner.
  - Overall data element agreement rate for all sites was 90.4 percent.
- Empirical validity testing conducted at the Accountable Entity Level:
  - Sensitivity was 100% in all reliability pilot sites and specificity was 100 percent in pilot sites 1, 2, 3, 6, and 7 and 62.5 percent in pilot site 9 and 90.48 percent across pilot sites. The overall positive predictive value was 94.7 percent and negative predictive value was 100 percent in all pilot sites.
  - Overall measure outcome agreement rate was 91.2 percent with a kappa score of 0.881.
- Face validity testing conducted at the Accountable Entity Level:
  - A Technical Expert Panel (TEP) consisting of 15 members was convened, of which 80% of the members agreed and 20% moderately agreed that this is an important health outcome measure because there is room for improvement. The majority of members agreed that this rate is a critical component of defining and comparing quality of obstetric care between hospitals.
  - A patient working group consisting of five members strongly agreed that the measure is important because there is room for improvement and that it can be used to differentiate quality in obstetric outcomes between hospitals.
- The Feasibility Scorecard indicated that the following data elements have issues with accuracy:
  - Laboratory Test, Performed, Result dateTime PaO2/FiO2
  - Laboratory Test, Performed, Result PaO2/FiO2
  - Encounter Performed, Diagnosis, Present On Admission Indicator
  - Procedure, Performed, Relevant Period stopTime (Conversion of Cardiac Rhythm)
  - Procedure, Performed, Relevant Period stopTime (Hysterectomy)
  - Procedure, Performed, Relevant Period stopTime (Tracheostomy)
  - Procedure, Performed, Relevant Period stopTime (Delivery Procedures)
  - Procedure, Performed, Relevant Period stopTime (Ventilation)

Exclusions

- Patients with a confirmed diagnosis of COVID with COVID-related respiratory condition and patients with confirmed diagnosis of COVID with COVID-related respiratory procedure were excluded from the measure. Available studies suggest that symptomatic pregnant women

with COVID-19 are at increased risk of more severe illness compared with nonpregnant peers so the developer added this exclusion to ensure patients with this condition who were symptomatic with respiratory conditions would not be counted as a numerator case for hospitals.

- A total of 0.06% of denominator cases were excluded due to the COVID criteria (n=37). The range of denominator exclusions was from 0 to 6 cases per hospital. The developer states this had minimal impact on the performance scores and that since the number of pilot sites was so small they did not perform any formal statistical tests on this data.

Risk-Adjustment

- A hierarchical logistic regression risk model (HLM) was developed for both severe obstetric complications and severe obstetric complications excluding blood transfusion-only encounters. HLM accounts for hospital level clustering. This includes a random intercept for the hospital-specific effect. The model was tested using data from eight pilot sites, the data from which were divided in a 70/30 split for a developmental and validation dataset.
- The risk model includes demographics and patient characteristics, pre-existing conditions and pregnancy characteristics, lab tests and vital signs upon hospital arrival, long-term anticoagulant medication use, and a social risk factor for economic/housing instability.
- Risk variables were removed from inclusion in the model if there were greater than 20% missing values which were relevant for vital signs and lab results.
- The calculated C-statistic for the risk model for any severe obstetric complications was 0.74 using the development dataset and 0.75 using the validation dataset; the calculated C-statistic for the severe obstetric complications excluding blood transfusion-only cases measure was 0.77 using the development dataset and 0.73 using the validation dataset.
    - The calibration indices (γ0, γ1) used to assess the risk model for the any severe obstetric complications in the validation dataset are (0.15, 1.05) and for the severe obstetric complications excluding blood transfusion-only cases in the validation dataset are (0.22, 1.04). The calibration values which are consistently close to 0 at one end and close to 1 at the other end indicate good calibration of the model.
    - The two predictive models had an area under the ROC curve of 0.74 and 0.77 for any severe obstetric complications and severe obstetric complications excluding blood transfusion only cases, respectively. This moderate level of predictive ability demonstrates that controlling for these identified patient characteristics in measure calculations should control for differences in patient characteristics across hospitals.
    - Model fit was also assessed using model Chi-square which shows the models are significantly better than the null models.
- Housing/economic instability was included as a risk factor and race/ethnicity as a stratification factor. These decisions were made *a priori* and were not tested or influenced by analytic results.
    - Because of the stark differences in maternal outcomes by race/ethnicity as demonstrated in the literature, these social risk factors were examined as stratification variables rather than risk variables. It was determined that illumination of outcome disparities by race/ethnicity, rather than adjustment of outcomes by race/ethnicity, would best inform stakeholders and patients and be most impactful in

incentivizing improvements in quality of maternal care. These results were not included in the submission.

Meaningful Differences

- The developer shows variation in pilot site severe obstetric complication rates and states that this indicates a clinically meaningful quality gap in the delivery of maternal care to patients experiencing a delivery hospitalization.
- For the outcome of any severe obstetric complications:
    - Pilot site risk standardized results ranged from 158 delivery encounters with severe obstetric complications to 299 delivery encounters with severe obstetric complications.
    - Pilot hospital risk standardized results ranged from 157 delivery encounters with severe obstetric complications to 369 delivery encounters with severe obstetric complications
- For the outcome of severe obstetric complications excluding blood transfusion-only cases:
    - Pilot site risk standardized results ranged from 48 to 55 delivery encounters with severe obstetric complications.
    - Pilot hospital risk standardized results ranged from 49 to 55 delivery encounters with severe obstetric complications.

Missing Data

- Many of the data elements used in the Severe Obstetric Complications eCQM are defined with ICD-10 diagnosis or procedure codes (for example, severe maternal mortality numerator events and risk adjustment variables). The developer states that none of these data elements are considered to be missing when absent since the absence of a given code implies absence of the corresponding condition.
- For data elements representing vital signs and lab results, it is clinically acceptable that certain vital signs and labs were not performed for certain patients. However, vital sign and lab result fields with more than 20% missing were not considered as potential risk adjustment variables based on statistical considerations.
- Two pilot sites had mismatches due to missing data. Pilot Site 1 had only one case resulting in a mismatched measure outcome. The ICD-10 delivery code was missing from the procedure list and therefore the patient did not land in the initial population based on extracted data but in the denominator based on the adjudicated data. Pilot Site 3 had one of the cases mismatched based on a missing delivery time. This error resulted in the patient qualifying for the initial population based on the original data and qualifying for the denominator based on the adjudicated data.

SMP Summary:

- The validity testing approaches were largely seen as acceptable; however, a concern was raised that the face validity testing lacked testing of the exclusion for COVID and the 34 risk adjustment variables.
- Data element validity testing was incomplete because not all elements were tested.

- The risk adjustment methodology was seen as appropriate but there were questions about how stratification by social factors (i.e. race and housing insecurity) may work in real world use.

Questions for the Committee regarding validity:

- *Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?*
- *Does the absence of race/ethnicity stratification results cause any concern?*
- *Are the accuracy issues captured in the Feasibility Scorecard substantial enough to impact the validity of these data elements?*
- *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?*
- *Are the accuracy issues captured in the Feasibility Scorecard substantial enough to impact the validity of these data elements?*

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

Committee Pre-evaluation Comments:

2a. Reliability

- 2a1. Reliability-Specifications
  - Blood transfusions have been eliminated from the specifications used in federal surveillance (HRSA, AHRQ, CDC page not yet update) due to poor positive predictive value in the absence of other SMM indicators
  - No concerns
  - No concerns
  - With the electronic design of the measure, consistent implementation should occur.
- 2a2. Reliability – Testing
  - It is concerning that reliability dips when excluding blood transfusion only cases when these should be excluded due to poor predictive value
  - No concerns
  - No concerns
  - No

2b. Validity

- Chart review for documentation of the 21 indicators is not the same as comparison to gold standard criteria for true SMM established by ACOG and medical experts. Validation studies in California, Pittsburgh, Boston, and Ann Arbor have shown poor predictive value with as many as half of all cases identified with these codes not being truly severe. We also know that state-level variation in SMM is inconsistent with variation in other perinatal indicators, including maternal mortality. If this indicator is not comparable across states, how can it be comparable across hospitals and used to assess health care quality and performance? There is an urgent need for measure refinement. It is not ready for primetime as a comparative measure of hospital performance. Please see the following in the ACOG consensus statement: "Definitions of severe maternal morbidity that rely on diagnosis codes, such as the Centers for Disease Control and Prevention's definition, may miss cases, have a relatively low

positive predictive value (0.40) and, at a practical level, may be difficult for facilities to operationalize 10. Facilities should have a screening process in place to detect cases of severe maternal morbidity for review. The College and SMFM recommend using two criteria to screen for severe maternal morbidity: 1) transfusion of 4 or more units of blood and 2) admission of a pregnant or postpartum woman to an ICU. Investigators have demonstrated that these criteria have high sensitivity and specificity for identifying women with severe morbidity and a high positive predictive value (0.85) for identifying severe maternal morbidity" Why isn't a more simple accurate measure of 4+ units of blood products or ICU admission being used as the outcome measure or to validate?

- No concerns
- Concerns re: race & ethnicity data. See 2b2-3.
- No

2b2-2b6. Potential threats to validity

- 2b2-3. Other Threats to Validity (Exclusions, Risk Adjustment)
    - Risk adjustment is necessary to account for pre-existing health status. It appears to be appropriate.
    - All risk-adjustment variables are present at the start of care; models have good c-statistic
    - I do not understand why this measure is not or cannot be stratified for race & ethnicity. These are important data points and should be collected and reported. I would like to hear from the developer why only housing insecurity was chosen as a risk factor.
    - No concerns
- 2b4-7. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data)
    - Meaningful differences in care quality are not established without comparison to gold standard criteria or restriction to indicators with greater preventability https://www.acog.org/clinical/clinical-guidance/obstetric-care-consensus/articles/2016/09/severe-maternal-morbidity-screening-and-review
    - No concerns - missing data not a problem; saw variation in calculated rates across entities
    - No concerns
    - There are significant differences in performance noted among delivering facilities.

Criterion 3. Feasibility

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

- All data elements included in the measure score are generated or collected during the provision of care, and all data are in defined fields in a combination of electronic sources. The information is coded by someone other than the person obtaining the original information.]
- Using a simulated data set, the submission demonstrates that the evaluation of 100% of the measure logic can be automated.
- The Feasibility Scorecard assesses each data element across the following domains:
    - Availability - is the data element readily available in a structured format across EHR systems?
    - Accuracy- is the information contained in the data is correct?
    - Standards - is the data element coded using a nationally accepted terminology standard?
    - Workflow - is the data element routinely captured and used during care delivery?

- The developer has identified feasibility issues for the following data elements. For each data element the developer was asked to provide additional context for the issue and a plan for addressing the issue.
    - Laboratory Test, Performed, Result dateTime PaO2/FiO2
    - Laboratory Test, Performed, Result PaO2/FiO2
    - Encounter Performed, Diagnosis, Present On Admission Indicator
    - Procedure, Performed, Relevant Period stopTime (Conversion of Cardiac Rhythm)
    - Procedure, Performed, Relevant Period stopTime (Hysterectomy)
    - Procedure, Performed, Relevant Period stopTime (Tracheostomy)
    - Procedure, Performed, Relevant Period stopTime (Delivery Procedures)
    - Procedure, Performed, Relevant Period stopTime (Ventilation)
- The developer determined several of the pilot sites were unable to accurately capture two main data elements: the timestamp for the procedure performed and the laboratory test result of the Pa02/Fi02 ratio. The draft specifications were revised to better align with clinical intent and decrease burden for a lab result that was not commonly calculated in the EHR. Feasibility scores based on the revised specifications increased to 98%.

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?*
- *For data elements assessed to have feasibility issues, does the developer present a credible, near-term path to electronic collection?*

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

- Moderate – all identified feasibility issues have a core plan to address the issues and 100% coverage in simulated data unit tests (BONNIE)

Committee Pre-evaluation Comments:

3. Feasibility

- This is a complicated measure with 21 indicators and many other risk-adjustment factors raising concerns about real-world feasibility
- All data elements included in the measure score are generated or collected during the provision of care, and all data are in defined fields in a combination of electronic sources; 100% of the measure logic can be automated
- No concerns.
- No concerns

Criterion 4: Use and Usability

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

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

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

Current uses of the measure

Publicly reported? ☒ Yes ☐ No

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

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

Accountability program details

- This measure is used in the ORYX Performance Measure Reporting: Hospital Accreditation Program (HAP) and Critical Access Hospital Accreditation (CAH) Program, implemented by The Joint Commission.
- These programs also provide quality improvement data with both internal and external benchmarking. The data submitted is analyzed by The Joint Commission for trends and benchmarks and for internal quality improvement purposes.

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

- After the pilot testing concluded and final results were analyzed, a pilot summary report was created and shared with each pilot site via email.
- The Joint Commission developed dashboards as part of the ongoing continuous customer engagement project. The dashboard report, posted in the Resources and Tools section of an accredited hospital's secure Joint Commission Connect® extranet site, is representative of each organization's relative performance on each of the selected measures.
- Feedback was obtained during a public comment period for those being measured. Commenters provided support for focusing measurement on addressing severe maternal morbidity and improving maternal health outcomes, the usefulness of this measure in assessing and improving the quality of care for patients, and publicly reporting both an overall rate of severe obstetric complications and a rate of severe obstetric complications excluding blood transfusion-only cases.
- Feedback was obtained from a TEP and patient working group. Experts and patients expressed that this is an important health outcome measure with room for improvement and it would distinguish between hospital performance.

***Questions for the Committee**:*

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

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

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

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

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

Improvement results

- As this is a new measure, performance improvement data is not yet available.

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

Unexpected findings (positive or negative) during implementation

- There are no implementation findings at this time.

Potential harms

- The developer notes that measuring obstetric complications may cause a shift in hospital resources to support EHR data extraction and reporting and away from other functions. Also, hospitals may potentially focus on complications in the measure while dismissing other complications not currently measured.

Additional Feedback:

- This measure was reviewed by the Measure Applications Partnership (MAP) for the Interoperability and Inpatient Quality Reporting (IQR) programs in 2021. The MAP recommended conditional support for rulemaking in both programs pending the successful completion of testing and CBE endorsement.
    - The MAP's rationale for conditional support in the interoperability program was that this measure provides meaningful use of certified electronic health record technology and that it would be the only measure in the Interoperability program addressing maternal health and obstetric complications if included.

- MAP's rational for this measure's conditional support in the IQR program was it would be the only outcome measure in Hospital IQR that directly measures morbidity and obstetric complications.

- For both programs, MAP raised concerns about the sample size used for testing in the measure.

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

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

Committee Pre-evaluation Comments:

4a. Use

- It doesn't seem like feedback from medical experts was really received or integrated. Please see ACOG consensus statement that was reaffirmed in 2021 https://www.acog.org/clinical/clinical-guidance/obstetric-care-consensus/articles/2016/09/severe-maternal-morbidity-screening-and-review
- Feedback was given via report and online dashboard; public comment from those being measured expressed support
- No concerns. Should be required of all birthing facilities.
- Yes

4a. Usability

- While a measure of severe obstetric complications is needed, the current specifications do not accurately capture severe morbidity and may reflect coding practices more than care quality. Thus, the current specification could do real harm in not actually assessing true morbidity that's reflective of care quality.
- No concerns.
- No concerns
- None

Criterion 5: Related and Competing Measures

- No related or competing measures identified.

Committee Pre-evaluation Comments:

5: Related and Competing Measures

- No related/competing measures.
- None

- Not that I'm aware of

Public and NQF Member Comments (Submitted as of Month Day, Year)

Member Expression of Support

- Of the X NQF members who have submitted a expression of support, X expressed "support" and X expressed "do not support" for the measure.

Comments

[Insert MIMS pre-evaluation comments export]

Scientific Acceptability Evaluation

RELIABILITY: SPECIFICATIONS

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

   Submission document: Items sp.01-sp.30

2. **Briefly summarize any changes to the measure specifications and/or concerns about the measure specifications.**

   Reviewer 6: none

   Reviewer 7: Complex and at times very hard to follow.

   Reviewer 9: No concerns.

   Reviewer 10: None

RELIABILITY: TESTING

3. **Reliability testing level:** ☒ **Accountable-Entity Level** ☒ **Patient/Encounter Level** ☐ **Neither**
4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure:**

   ☐ **Yes** ☐ **No**

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

   ☐ **Yes** ☐ **No**

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

   Submission document: Question 2a.10

   Reviewer 1: Accountable Entity Level ("Measure Score") Reliability: Measure scores were calculated for 8 pilot sites used for risk model development, and for the 25 individual hospitals within those 8 pilot sites. Measure score reliability was evaluated using a signal-to-noise ratio to assess the values according to conventional standards (Landis & Koch, 1977). For measure score reliability testing of measure scores for the 25 individual hospitals, testing was conducted at several volume thresholds, including: no required minimum number of delivery encounters for the year, at least 25 delivery encounters for the year, and 200 delivery encounters for the year.

   Reviewer 3: Data element: comparison of EHR abstracted data with chart review Facility: S/N from hierarchical model

Reviewer 5: acceptable

Reviewer 6: Reliability testing was performed using SNR analysis.

Reviewer 7: Appropriate

Reviewer 9: Signal to noise and agreement tests were done.

Reviewer 10: STN was used, approach is acceptable.

Reviewer 11: signal to noise ratio

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

   Submission document: Question 2a.11

   Reviewer 1: Results at the health site level (N=8 pilot sites) yielded a median reliability score of 0.991 (range: 0.982 –0.997) for any severe obstetric complication and 0.955 (range: 0.916 – 0.983) for severe obstetric complications excluding blood transfusion-only cases. Each pilot site had at least 25 delivery encounters.
   Signal-to-noise reliability was calculated for the 25 individual hospitals at several volume thresholds; results are provided for hospitals with at least 25 delivery encounters in the year (all hospitals included in testing) and for hospitals with at least 200 delivery encounters in the year (23 of the 25 hospitals included in testing). For hospitals with at least 25 delivery encounters, the median reliability score was 0.959 (0.802-0.996) for any severe obstetric complication outcome and 0.684 (0.273-0.961) for severe obstetric complications excluding blood transfusion-only cases. The signal-to-noise reliability is higher when included hospitals had at least 200 delivery encounters for the year, rather than 25 delivery encounters, particularly for the second outcome (severe complications excluding blood transfusion-only cases: the median reliability score was 0.978 (0.867-0.996) for any severe obstetric complication outcome and 0.804 (0.377-0.961) for severe obstetric complications excluding blood transfusion-only cases.

   Reviewer 2: Signal-to-noise reliability results at the hospital level indicate very high reliability for the outcome measuring any severe obstetric complication, and a lower reliability for the outcome measuring any severe obstetric complications excluding blood transfusion.

   Reviewer 3: Data element: Good. Specifications modified during development in response to difficulty obtaining selected measures. Facility: Reliability high on average for facilities with >200 births. Reliability for least reliable in this group below 0.6.

   Reviewer 5: acceptable

   Reviewer 6: Median SNR was 0.95 for severe OB complications and 0.68 for severe OB complications excluding blood transfusions.

   Reviewer 7: Signal to noise reliability at health site level is moderate to high.

   Reviewer 9: There seems to be a problem when blood transfusion cases are excluded.

   Reviewer 10: Results were mixed. Median reliability of 0.959 oberved for any severe obstetric complication, but reliability excluding blood transfusion only was 0.684 (IQR of 0.46 – 0.91), for hospitals with >25 encounters. This threshold minimum is not part of the specification, so it's possible the reliability is worse if all hospitals are included.

   Reviewer 11: high reliability

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: Question 2a.10-12

   ☒ **Yes** ☐ **No** ☒ **Not applicable**

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

   Submission document: Question 2a.10-12

☒ **Yes** ☐ **No** ☒ **Not applicable** (patient/encounter level 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 accountable-entity level testing has been conducted)

    ☒ **Moderate** (NOTE: Moderate is the highest eligible rating if accountable-entity 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.**

    Reviewer 1: The signal-to-noise reliability results at the health site level show very high reliability for both outcomes.

    Reviewer 2: I'm unclear which outcome, level of aggregation, and minimum sample size I should be basing my judgements on. Site-level reliability is excellent. Hospital-level reliability is excellent for any complication but lower when transfusion-only cases are excluded (25% have reliability< 0.46). This improves when minimum sample is 200.

    Reviewer 3: S/N acceptable at 200 births but some unreliability issues at low end for measure if transfusion-only cases excluded.

    Reviewer 4: The developers don't provide the full technical details of reliability estimation. The results are encouraging assuming the methods are appropriate.

    Reviewer 6: Median SNR was 0.95 for severe OB complications and 0.68 for severe OB complications excluding blood transfusions.

    Reviewer 7: No concerns.

    Reviewer 10: Developer used a threshold that is not part of the specification. Reliability for complications excluding blood transfusions was marginal.

VALIDITY: TESTING

12. **Validity testing level (check all that apply):**

    ☐ **Accountable-Entity Level** ☒ **Patient or Encounter-Level** ☒ **Both**

13. **If patient/encounter level validity testing was provided, was the method described and appropriate for assessing the accuracy of ALL critical data elements? NOTE:** Data element validation from the literature is acceptable.

    ☐ **Yes**

    ☐ **No**

    ☒ **Not applicable** (patient/encounter level testing was not performed)

14. **Method of establishing validity at the accountable-entity level:**

    ☐ **Face validity**

    ☐ **Empirical validity testing at the accountable-entity level**

    ☒ **N/A (accountable-entity level testing not conducted)**

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

    Submission document: Question 2b.02

☒ **Yes**

☒ **No**

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

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

Submission document: Question 2b.02

Reviewer 1: A statistically representative sample of the electronically submitted inpatient encounters was selected for re-abstraction. During the virtual visits, site staff shared their screen, navigated through the electronic health records of the sampled patients while developer manually re-abstracted each data element. To determine validity, re-abstraction findings were compared with the original electronic data submission and any disagreements were adjudicated with reasons for discrepancies noted. The performance measure outcome rates were compared, and agreement rates were corrected for chance variation with the kappa statistic.

To assess face validity, a Qualtrics survey was produced and distributed to the members of the Technical Expert Panel (TEP) for their completion. A Likert scale was used with the 6 possible responses ranging from Strongly Agree to Strongly Disagree.

Reviewer 2: Item-level empirical testing and score level face validity.

Reviewer 3: Data element: accuracy of abstracted data Score: Face validity from TEP. No empirical validity tests.

Reviewer 5: acceptable

Reviewer 6: The results of the TEP analysis suggests that the method has face validity. The results of the assessment of the risk adjustment methodology (discrimination and calibration) suggests that the measure is valid at the accountable entity level.

Reviewer 7: Appropriate.

Reviewer 9: Methods appropriate

Reviewer 10: Face validity was appropriate but the data element approach lacked testing of the exclusion for COVID as well as the 34 risk adjustment variables.

Reviewer 11: compared with chart abstraction

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

Submission document: Questions 2b.03-04

Reviewer 1: Measure score validity testing was completed in the same pilot sites. The PPV (agreement rate) for the numerator among delivery encounters clinically adjudicated in validity testing was 100% at Pilot Sites 1, 2, 3, 6, and 7, and 70% at Pilot Site 9, with an overall PPV of 94.74%.
Specificity and sensitivity are high. Sensitivity is 100% in all reliability pilot sites and specificity is 100% in pilot sites 1, 2, 3, 6, and 7 and 62.5% in pilot site 9. This means that the probability of the EHR data detecting a true severe obstetric complication during a delivery hospitalization based on the abstracted data ('gold standard') is 100% (sensitivity). The probability of the EHR data accurately identifying that no severe obstetric complication occurred during a delivery hospitalization based on abstracted data ranged from 62.5% to 100% and was 90.48% across pilot sites (specificity). NPV was 100% in all pilot sites, indicating the EHR data indicated a severe obstetric complication did not occur, and 100% of the time the chart abstraction confirmed a severe obstetric complication did not occur.
15 members of the TEP completed face validity surveys. 80% of TEP members strongly agree while 20% moderately agree that this is an important health outcome to measure because there is room for improvement. 87% strongly or moderately agree the eCQM will produce reliable and valid rates while the

remaining 13% of respondents somewhat agree. Similarly, 87% strongly or moderately agree that hospitals can use the results for performance improvement, while the remaining 13% of respondents somewhat agree.

Reviewer 2: Item-level association between e-extracted data to chart review gold standard was very good. Face validity testing was generally supportive of the measure.

Reviewer 3: Data element: accuracy generally supported. Face validity: TEP supportive of measure, although some calls for additional validation. Some issues with risk adjustment, discussed below. One big issue on validity: Magnitude of score drops by 2/3 when transfusion-only events are excluded. Risk adjusted scores for testing sites for scores when transfusion-only events are excluded fit into a narrow range from 49 to 51/10000 births. I am concerned that virtually all the variation in risk adjusted scores are due to transfusion only events, and this needs to be discussed by the committee and the Standing Committee.

Reviewer 5: acceptable

Reviewer 6: data element validity for outcome data element: kappa = 0.88 (consistent with high level of validity) C statistic in validation data set are 0.75 and 0.73 for the 2 measures. Calibration indices are consistent with acceptable calibration.

Reviewer 7: Adequate.

Reviewer 9: There was an issue with site 9 which was addressed later in the application and specifications were changed. However, when this measure is used in practice, how will issues such as the ones site 9 had be identified? Will the site know they have erroneous rates and be able to take action?

Reviewer 10: The measure has evidence of face validity, data element validity testing was incomplete.

Reviewer 11: generally high agreement and kappa scores across all testing sites

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

Submission document: Questions 2b.15-18.

Reviewer 2: "denominator exclusion for COVID plus respiratory conditions was added post pilot due to the growing evidence of perinatal complications in women who have COVID infection with respiratory conditions." Why an exclusion instead of including it in the risk model?

Reviewer 3: Covid exclusion seems reasonable.

Reviewer 6: none

Reviewer 7: No concerns. Small number of exclusions.

Reviewer 9: No concerns.

Reviewer 10: No concerns

Reviewer 11: doesn't mention issue of transfer from birthing centers -- hopefully POA will filter out excludes only covid pulmonary diagnoses but covid may have other manifestations impacting obstetric complications

19. **Risk Adjustment**

19a. **Risk-adjustment method**

☐ None (only answer Question 20b and 20e) ☒ Statistical model ☒ Stratification

☐ Other method assessing risk factors (please specify)

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

☐ Yes ☐ No ☒ Not applicable

19c. **Social risk adjustment:**

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

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

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

19d. **Risk adjustment summary:**

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

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

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

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

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

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

Reviewer 3: Risk adjustment approach uses clinical present-on-admission comorbidities, which appear to have a substantial association with the risk of severe obstetrical complications. The risk adjustment model has one social determinant: economic/housing instability. And in the models, this variable has an OR of 1.79, sizable but not an outlier, in the overall model, but an OR of 5.10 in the risk adjustment model that excludes transfusion only cases, which is the largest OR in this model. I would like to know much more about how this variable is obtained and coded and its reliability, and discussion of whether it should be included in the risk adjustment model. In the risk adjustment table, only 62 cases 0.01% are reported as having economic/housing instability. This seems too low, another reason for concern. The developers propose stratifying the analysis by race/ethnicity but indicate the approach is still under development. The committee should discuss.

Reviewer 7: Appropriate. Continue working on race/ethnicity stratification for this measure.

Reviewer 9: No concerns

Reviewer 10: approach is appropriate.

Reviewer 11: sponsors have elected not to include social risk factors in the risk adjustment modeling but in stratification in subsequent reporting--exactly how this will play out is not clear--i.e. will hospital scores be compared only within various strata of race and housing insecurity?

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

For cost/resource use measures, does this measure identify meaningful differences about cost and resource use between the measured entities?

Submission document: Questions 2b.05-07

Reviewer 2: As shown in Table 2b.06.02, there is virtually no variability in risk adjusted outcomes with transfusion-only cases are excluded.

Reviewer 3: As noted above, the range of risk adjusted rates/10000 when transfusion-only cases are excluded narrows to 49-51. The range before risk adjustment is broader.

Reviewer 6: none

Reviewer 7: No concerns.

Reviewer 9: No concerns

Reviewer 10: No concerns.

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

Submission document: Questions 2b.11-14.

Reviewer 3: NA
Reviewer 6: none
Reviewer 9: N/A
Reviewer 10: n/a

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

    Submission document: Questions 2b.08-10.

    Reviewer 3: NA. Adjustments made during development in response to identified data problems.

    Reviewer 6: none

    Reviewer 7: No concerns.

    Reviewer 9: The measure specifications were changed to address this.

    Reviewer 10: no concerns.

    Reviewer 11: Missing data is addressed somewhat appropriately

For cost/resource use measures ONLY:

If not cost/resource use measure, please skip to question 25.

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

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

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

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

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

    ☒ **Moderate** (NOTE: Moderate is the highest eligible rating if accountable-entity 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 accountable-entity level and the patient/encounter level is required; if not conducted, should rate as INSUFFICIENT.)

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

    Reviewer 1: Strong PPV and NPV results with high specificity and sensitivity. Overall, the study revealed ePC-07 to have an excellent measure outcome agreement rate of 91.2% with a kappa score of 0.881 indicating strong agreement. Well developed statistical risk adjustment model with 1 social risk factor (housing stability) and stratification by race/ethnicity. It is well known that maternal outcomes among Black Americans are worse and thus stratification seems acceptable to highlight these disparities. Model statistics were strong. The calculated C-statistics of 0.74 and 0.75 for the risk model for any severe obstetric complications (development and validation datasets), and 0.77 and 0.73 for the severe obstetric complications excluding blood transfusion-only cases measure (development and validation datasets), indicate good model discrimination.

    Reviewer 2: I'm unclear which outcome, level of aggregation, and minimum sample size I should be basing my judgements on. As shown in Table 2b.06.02, there is virtually no variability in risk adjusted outcomes with transfusion-only cases are excluded. My rating of moderate is based on the outcome of Any complication.

Reviewer 3: I need more information on the narrow range of risk adjusted scores for transfusion-only events and discussion of the apparent outsize role transfusion-only events play in the variation in scores across hospitals.

Reviewer 6: The performance of the risk adjustment model is acceptable.

Reviewer 10: As the developer notes, The face validity assessment demonstrated that the Technical Expert Panel members believe this eCQM is an important health outcome to measure because there is room for improvement, it will produce reliable and valid rates, and hospitals can use the results for performance improvement.

Reviewer 11: Data presented suggests a very high rating for validity--clarification of how stratification for social factors will be accomplished is needed

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

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

    ☐ **High**

    ☐ **Moderate**

    ☐ **Low**

    ☐ **Insufficient**

28. **Briefly explain rationale for rating of EMPIRICAL ANALYSES TO SUPPORT COMPOSITE CONSTRUCTION**
    - [Summary]

ADDITIONAL RECOMMENDATIONS

29. **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.**
    - [Summary]

Developer Submission

Criteria 1: 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

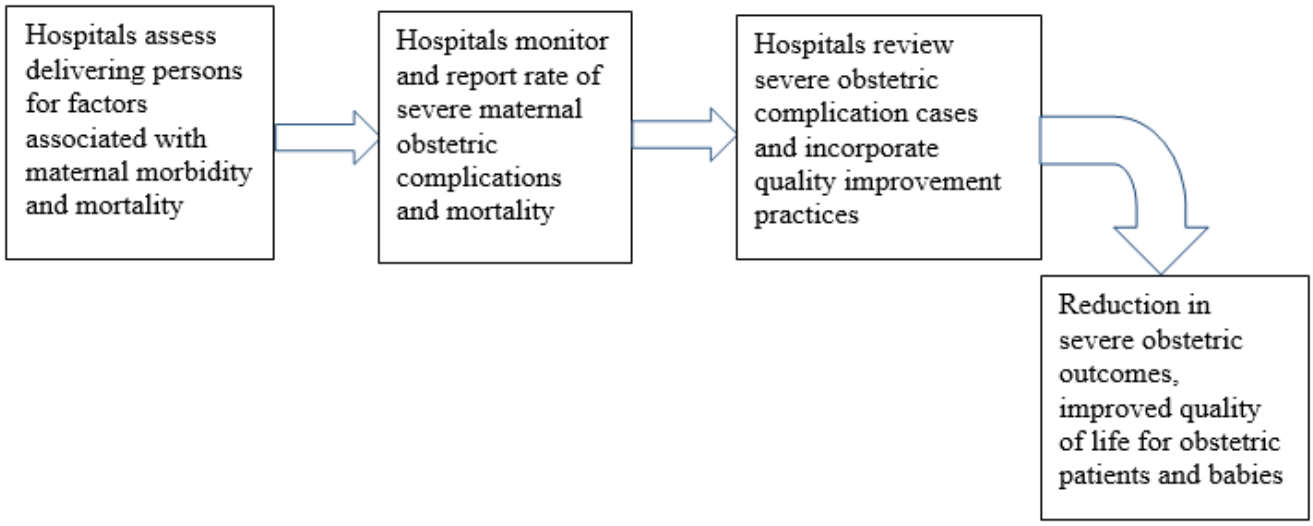
---

Please separate added or updated information from the most recent measure evaluation within each question response in the Importance to Measure and Report: Evidence section. For example:
2021 Submission:
Updated evidence information here.
2018 Submission:
Evidence from the previous submission here.

1a.01. Provide a logic model.

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.

[Response Begins]



The goal for this measure is to assess the occurrence of specific severe obstetric complications in the hospital setting by using a methodology that reliably allows comparison across hospitals. Reduction in maternal complications will reduce maternal death and disability and improve maternal quality of life. The Severe Obstetric Complication electronic clinical quality measure (eCQM) is expected to inform hospital efforts to improve maternal health outcomes and thus reduce the costs associated with adverse health outcomes. The measure specifications are harmonized with other perinatal measures (for cohort alignment) and with the Center for Disease Control and Prevention's (CDC's) 21 indicators of severe maternal morbidity (SMM) (for harmonization of the measure outcome) for broad applicability across hospitals.
[1]Geller SE, Rosenberg D, Cox SM, et al. The continuum of maternal morbidity and mortality: factors associated with severity. *American journal of obstetrics and gynecology.* 2004;191(3):939-944.

[Response Ends]

1a.02. Provide evidence that the target population values the measured outcome, process, or structure and finds it meaningful.

Describe how and from whom input was obtained.

[Response Begins]
To gain targeted input from the patient and caregiver perspective, a Patient Working Group was recruited through collaboration with Rainmakers Strategic Solutions LLC. The Patient Working Group was composed of seven members, including patients and caregivers with diverse experiences and perspectives. The first Patient Working Group meeting was held in August 2020 via web-based webinar during which Patient Working Group members provided input on initial measure specifications for the measure cohort, outcome and risk adjustment. The second meeting was held in July 2021 via web-based webinar, at which Patient Working Group members provided input on measure specification updates, as well as feasibility testing and reliability results and initial validity testing results. At the third meeting, a web-based webinar held in November 2021, Patient Working Group members provided input on the risk adjustment model, measure scores, and further testing results.
The Working Group members provided personal and insightful perspectives on key measure aspects of measure development and decisions. The members strongly believe this eCQM is an important health outcome to measure because there is room for improvement and strongly/moderately agree that this measure is a critical component of defining and comparing the quality of obstetric care between hospitals. See Section 2b.03 for further details on face validity.

[Response Ends]

1a.03. Provide empirical data demonstrating the relationship between the outcome (or PRO) and at least one healthcare structure, process, intervention, or service.

[Response Begins]
The high maternal mortality and morbidity rates in the United States present unique opportunities for large-scale quality measurement and improvement activities. Statistics on preventability vary but suggest that a considerable proportion of maternal mortality and morbidity events could be prevented. A 2019 report from 14 maternal mortality review committees conducting a thorough review of pregnancy-related deaths determined that 65.8% of deaths were preventable (Data from 14 U.S. Maternal Mortality Review Committees, 2008-2017).1 Additionally, a study that examined preventability of pregnancy-related death, women with near-miss morbidity, and those with severe morbidity found that 40.5% of deaths, 45.5% of near miss morbidity, and 16.7% of other severe morbidities were preventable.2 Geller et. al. identified areas of focus for preventability of morbidity and mortality included assessment/point of entry to care, diagnosis and recognition of high risk, referral to experts, treatment, management hierarchy, education, communication, policies and procedures, documentation, and discharge.
[1]Davis NL, Smoots AN, Goodman DA. Pregnancy-Related Deaths: Data from 14 US Maternal Mortality Review Committees. *Education.* 2019;40(36):8.2.
[2]Geller SE, Rosenberg D, Cox SM, et al. The continuum of maternal morbidity and mortality: factors associated with severity. *American journal of obstetrics and gynecology.* 2004;191(3):939-944.

[Response Ends]

1b.01. Briefly explain the rationale for this measure.

Explain how the measure will improve the quality of care, and list the benefits or improvements in quality envisioned by use of this measure.

[Response Begins]
The United States experiences higher rates of maternal morbidity and mortality than most other developed countries. These rates have continued to trend upward in recent decades.1 Research indicates that the overall rate of severe maternal morbidity (SMM) has increased by almost 200% between 1993 and 2014 to 144 per 10,000 delivery hospitalizations1, with more than 25,000 women per year experiencing obstetric complications.2 Recent maternal mortality data from 2018 reveal that 658 women died from maternal causes, resulting in a rate of 17.4 deaths per 100,000 live births, with 77% of the deaths attributed to direct obstetric causes like hemorrhage, preeclampsia, obstetric embolism, and other complications.3 This has prompted national health experts and organizations to prioritize quality improvement strategies to mitigate risk of adverse outcomes among maternal populations. The U.S. Department of Health & Human Services (HHS) has also called for action to improve maternal health and outcomes and outlines seven

actions for healthcare professionals, including participating in quality improvement and safety initiatives.4 There are currently only a small number of quality measures focused on maternal health, and those implemented at the national level are mostly process measures and limited in scope. While these existing measures aim to promote coordination of care and standardize health care processes, maternal health outcome measures are sorely needed. Measures that are focused on maternal health outcomes will address the patient safety priority area under the Meaningful Measures 2.0 framework, and likewise will use EHR data to address interoperability, another meaningful measure area for assessing quality of health care.5

1. Centers for Disease Control and Prevention. Severe Maternal Morbidity in the United States. January 31, 2020; https://www.cdc.gov/reproductivehealth/maternalinfanthealth/severematernalmorbidity.html.

[Response Ends]

1b.02. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.

Include mean, std dev, min, max, interquartile range, and 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 (4b) under Usability and Use.

[Response Begins]

There are a limited number of pilot hospitals, therefore the five number statistical summaries are used in place of the scores by deciles. Data for 25 hospitals are summarized in Table 1b.02 at the hospital level for 2020 discharges using a rate per 10,000 deliveries. Maternal morbidity data in literature is reported as rates per 10,000 and maternal mortality rates are reported per 100,000. The Severe Obstetric Complications rate includes both maternal morbidity and mortality and is reported as a rate per 10,000. The median number of encounters was 799 per hospital site.

Table 1b.02.01 Risk- adjusted Hospital Level Rates

Statistic	Value
Mean	248.8, *SD 55.5*
Min	157.1
25th Percentile	215.6
50th Percentile	238.2
75th Percentile	287.3
Max	369.5

Risk-adjusted rates per 10,000 on this measure

Table 1b.02.01 displays the statistical measurements of the risk-adjusted hospital level rates. See above paragraph for specific details.

For reference, each health system will be referred to as a 'pilot site' and 'hospital' will refer to the individual hospitals within the health system. A total of 10 pilot sites, consisting of 28 hospitals were included in some phase of pilot testing.

Table 1b.02.02 Pilot Site Characteristics

Pilot Site ID	# of Hospitals	State	Ownership Type*	Geography* (Urban, Suburban, Rural)	# Beds*	# of Births*	Teaching Program in OB/GYN*
Pilot Site 1	10	NC, VA	Nongovt. (not-for-profit) - Other	Urban	1807 (range 36 - 740)	16334 + (range 473 - 5568)	No
Pilot Site 2	1	RI	Nongovt. (not-for-profit) - Other	Urban	247	8823	No
Pilot Site 3	1	LA	Nongovt. (not-for-profit) - Other	Urban	228	8295	No
Pilot Site 4[a]	2	CA	Nongovt. (not-for-profit) - Church Operated	Urban	446	2921	No

Pilot Site ID	# of Hospitals	State	Ownership Type*	Geography* (Urban, Suburban, Rural)	# Beds*	# of Births*	Teaching Program in OB/GYN*
Pilot Site 5	9	OH, MI	Nongovt. (not-for-profit) – Other, Govt. (non federal) - County	6 Urban 3 Rural	1653 (range 35 - 595)	9283 + (range 165 - 3596)	No
Pilot Site 6	1	NJ	Nongovt. (not-for-profit) - Other	Urban	446	3319	No
Pilot Site 7	1	CA	Nongovt. (not-for-profit) - Other	Urban	541	4660	Yes
Pilot Site 8[b]	1	IL	Nongovt. (not-for-profit) - Other	Urban	650	2442	Yes
Pilot Site 9	1	MD	Nongovt. (not-for-profit) - Other	Urban	401	3854	No
Pilot Site 10[c]	1	PA	Nongovt. (not-for-profit) - Other	Urban	321	8796	Yes

* Source: American Hospital Association (AHA) DataQuery™ product, at the URL https://guide.prod.iam.aha.org/dataquery/reports, accessed March 16, 2021

a. Pilot Site 4 declined continued participation after Feasibility Testing

b. Data from Pilot Site 8 was collected but not available in time for Risk Model Development

c. Pilot Site 10 joined after Feasibility Testing

Table 1b.02.02 displays the characteristics of the entities measured. The information was retrieved from the American Hospital Association (AHA) DataQuery™ product. For each pilot site, the table provides the number of hospitals, the state, ownership type, whether the site is located in an urban, suburban or rural setting, the number of beds, births and if the hospital has a teaching program in obstetrics and gynecology.

[Response Ends]

1b.03. If no or limited performance data on the measure as specified is reported above, 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. Include citations.

[Response Begins]

The United States experiences higher rates of maternal morbidity and mortality than most other developed countries. These rates have continued to trend upward in recent decades.1 Research indicates that the overall rate of severe maternal morbidity (SMM) has increased by almost 200% between 1993 and 2014 to 144 per 10,000 delivery hospitalizations1, with more than 25,000 women per year experiencing obstetric complications.2 Recent maternal mortality data from 2018 reveal that 658 women died from maternal causes, resulting in a rate of 17.4 deaths per 100,000 live births, with 77% of the deaths attributed to direct obstetric causes like hemorrhage, preeclampsia, obstetric embolism, and other complications.3 This has prompted national health experts and organizations to prioritize quality improvement strategies to mitigate risk of adverse outcomes among maternal populations. The U.S. Department of Health & Human Services (HHS) has also called for action to improve maternal health and outcomes and outlines seven actions for healthcare professionals, including participating in quality improvement and safety initiatives.4 There are currently only a small number of quality measures focused on maternal health, and those implemented at the national level are mostly process measures and limited in scope. While these existing measures aim to promote coordination of care and standardize health care processes, maternal health outcome measures are sorely needed. Measures that are focused on maternal health outcomes will address the patient safety priority area under the Meaningful Measures 2.0

framework, and likewise will use EHR data to address interoperability, another meaningful measure area for assessing quality of health care.5

1. Centers for Disease Control and Prevention. Severe Maternal Morbidity in the United States. January 31, 2020; https://www.cdc.gov/reproductivehealth/maternalinfanthealth/severematernalmorbidity.html.
2. U.S. Department of Health & Human Services. HHS Outlines New Plans and a Partnership to Reduce U.S. Pregnancy-related Deaths. 2020; https://www.hhs.gov/about/news/2020/12/03/hhs-outlines-new-plans-to-reduce-us-pregnancy-related-deaths.html.
3. Hoyert DL, Miniño AM. Maternal mortality in the United States: changes in coding, publication, and data release, 2018. 2020.
4. U.S. Department of Health & Human Services. The Surgeon General's Call to Action to Improve Maternal Health. 2020.
5. Centers for Medicare & Medicaid Services. Meaningful Measures 2.0: Moving from Measure Reduction to Modernization. 2020; https://www.cms.gov/meaningful-measures-20-moving-measure-reduction-modernization, 2020.

[Response Ends]

1b.04. 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.

Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included. Include mean, std dev, min, max, interquartile range, and scores by decile. 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 (4b) under Usability and Use.

[Response Begins]

Risk ratios are provided in Table 1b.04.01 for race/ethnicity and payor. Age is not provided because it is included in the risk model. When adjusting for risk factors, Non-Hispanic - African American women have an 18% increased risk of having any SMM compared to non-Hispanic-white women, while Hispanic women had a 41% increased risk and Non-Hispanic-Asian/Pacific Islander women had a 62% increased risk for any SMM. When excluding blood transfusion only cases, compared to Non-Hispanic-White women, there was a 6% increased risk for Non-Hispanic-African American, 36% increased risk for Hispanic, and a 43% increased risk for Non-Hispanic-Asian/Pacific Islander women. When compared to private insurance, Medicaid and Medicare payors also showed an increased risk when adjusting for risk factors for any SMM and SMM excluding blood transfusion only cases.

Risk factor variables included in the risk adjustment model are as follows:

- Demographics and patient characteristics: maternal age
- Preexisting conditions and pregnancy characteristics defined by ICD-10 codes
- Anemia
- Asthma
- Autoimmune disease
- Bariatric surgery
- Bleeding disorder
- Body Mass Index (BMI)
- Cardiac disease
- Gastrointestinal disease
- Gestational diabetes
- Human Immunodeficiency Virus (HIV)
- Hypertension
- Mental health disorder
- Multiple pregnancy
- Neuromuscular disease
- Obstetric venous thromboembolism (VTE)
- Other pre-eclampsia
- Placental accreta spectrum
- Placental abruption
- Placenta previa
- Preexisting diabetes

- Preterm birth
- Previous cesarean
- Pulmonary hypertension
- Renal disease
- Severe pre-eclampsia
- Substance abuse
- Thyrotoxicosis
- Laboratory tests and vital signs upon hospital arrival (Hematocrit, White blood cell [WBC] count, Heart rate, Systolic blood pressure)
- Long-term anticoagulant medication use
- Social Risk Factors: economic/housing instability

Table 1b.04.01 represents data from 25 hospitals using 2020 discharges.

Table 1b.04.01 Race/Ethnicity and Payer Risk Adjustment Rate Ratios

Variable	Prevalence of risk factors n (%)	Any SMM Adjusted rate ratio (95% CI)	SMM excluding blood transfusion only cases Adjusted rate ratio (95% CI)
Race/Ethnicity	*	*	*
Non-Hispanic - White	33,371 (55.4%)	*	*
Declined/Unknown	1,916 (3.2%)	1.03 (0.75, 1.41)	1.23 (0.65, 2.30)
Hispanic	8,431 (14.0%)	1.41 (1.19, 1.67)	1.36 (0.95, 1.96)
Non-Hispanic - African American	11,853 (19.7%)	1.18 (1.02, 1.36)	1.06 (0.77, 1.47)
Non-Hispanic - Asian/Pacific Islander	2,932 (4.9%)	1.62 (1.26, 2.10)	1.43 (0.82, 2.49)
Non-Hispanic - Other	1,681 (2.8%)	1.15 (0.81, 1.63)	0.71 (0.28, 1.78)
Payer	*	*	*
Private Insurance	41,066 (68.2%)	*	*
Medicaid	16,221 (27.0%)	1.20 (1.05, 1.37)	1.13 (0.84, 1.50)
Medicare	223 (0.4%)	1.56 (0.87, 2.79)	1.47 (0.51, 4.24)
Other	2,518 (4.2%)	1.09 (0.82, 1.44)	0.89 (0.46, 1.72)
Self-pay or Uninsured	149 (0.2%)	0.47 (0.11, 1.98)	NA

NA: Not available due to small count

*Cells intentionally left blank

Table 1b.04.01 displays risk-adjustment rate ratios divided among race/ethnicity and between payers. The prevalence rate is provided and the rate ratio for any SMM and SMM excluding blood transfusion only cases.

Table 1b.04.02 Unadjusted Measure Rates per 10,000 by Age Category

Age	rate	n
<20	304.7	2363
20-25	300.3	9757
25-30	231.2	17259
30-35	214.3	20627
35-40	244.3	10847
40+	420.5	2307

Table 1b.04.02 displays the unadjusted measure rates per 10,000 for each age category. The n is also displayed by age category. The highest unadjusted measure rates are seen in the less than 20 and 40 plus age groups.

Table 1b.04.03 Unadjusted Measure Rates per 10,000 by Race Category

Race	rate	n
American Indian or Alaska Native	308.6	324
Asian	260.6	3108
Black or African American	353.1	14245

Race	rate	n
Native Hawaiian or Other Pacific Islander	331.1	151
Other Race	231.4	6266
White	210.1	38698
Patient Did Not Identify	310.9	193
Missing or Unknown	285.7	175

Table 1b.04.03 displays unadjusted measure rates per 10,000 for each race category. The n is also displayed by race category. The highest unadjusted rates are seen among the Black or African American and Native Hawaiian or Other Pacific Islander.

Table 1b.04.04 Unadjusted Measure Rates per 10,000 by Hispanic Ethnicity Category

Hispanic ethnicity	rate	n
Hispanic or Latino	257.8	8651
Not Hispanic or Latino	246.4	52770
Patient Did Not Identify	382.9	444
Unknown	340.6	411
Missing	158.4	884

Table 1b.04.04 displays unadjusted measure rates per 10,000 by Hispanic Ethnicity category. The n is also displayed by Hispanic Ethnicity category. The highest unadjusted rates are among patients who did not identify and unknown categories.

Table 1b.04.05 Unadjusted Measure Rates per 10,000 by Payer Category

Payer	rate	n
Private Insurance	208.9	41506
Medicaid	346.6	17888
Medicare	592.3	287
Other	223.4	3670
Self-pay or Uninsured	136.1	147

Table 1b.04.05 displays the unadjusted measure rates per 10,000 by payer category. The n is also displayed by payer category. The highest unadjusted rates are among Medicare and Medicaid payers.

[Response Ends]

1b.05. If no or limited data on disparities from the measure as specified is reported above, 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 above.

[Response Begins]

Our goal in selecting risk factors for adjustment was to develop parsimonious models that included clinically relevant variables strongly associated with a severe obstetric complication outcome. We used a two-stage approach, first identifying the comorbidity or clinical status risk factors that were most important in predicting the outcome, then considering the potential addition of social risk factors. Social risk factors considered were also dependent on the availability of information in the EHR. Economic/housing instability was included in the model and was chosen due to support in research literature for its inclusion and availability in the EHR. 1

Racial and ethnic disparities for women who identify as racial and ethnic minority groups are at a significantly higher risk for developing these complications than are Non-Hispanic White women.1 Because of the stark differences in maternal outcomes by race/ethnicity as demonstrated in the literature, these social risk factors were examined as stratification variables rather than risk variables, as discussed below. It was determined that illumination of outcome disparities by race/ethnicity, rather than adjustment of outcomes by race/ethnicity, would best inform stakeholders and patients and be most impactful in incentivizing improvements in quality of maternal care.

1. Leonard SA, Main EK, Scott KA, Profit J, Carmichael SL. Racial and ethnic disparities in severe maternal morbidity prevalence and trends. Annals of epidemiology. 2019; 33:30-36.

[Response Ends]

Criteria 2: 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.

---

sp.01. Provide the measure title.

Measure titles should be concise yet convey who and what is being measured (see What Good Looks Like).

[Response Begins]
ePC-07 Severe Obstetric Complications
[Response Ends]

sp.02. Provide a brief description of the measure.

Including type of score, measure focus, target population, timeframe, (e.g., Percentage of adult patients aged 18-75 years receiving one or more HbA1c tests per year).

[Response Begins]
Hospital-level measure scores are calculated as a risk-adjusted proportion of the number of delivery hospitalizations for women who experience a severe obstetric complication, as defined by the numerator, by the total number of delivery hospitalizations in the denominator during the measurement period. The hospital-level measure score will be reported as a rate per 10,000 delivery hospitalizations.
ePC07 was developed in collaboration with Yale New Haven Health Services Corporation/Center for Outcomes Research and Evaluation (CORE).

[Response Ends]

sp.04. Check all the clinical condition/topic areas that apply to your measure, below.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.
Please do not select:

- *Surgery: General*

[Response Begins]
Perinatal Health
Perinatal Health: Labor and Delivery
Perinatal Health: Post-Partum Care
[Response Ends]

sp.05. Check all the non-condition specific measure domain areas that apply to your measure, below.

[Response Begins]
Safety: Complications
[Response Ends]

sp.06. Select one or more target population categories.

Select only those target populations which can be stratified in the reporting of the measure's result.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.
Please do not select:

- *Populations at Risk: Populations at Risk*

[Response Begins]
Women
[Response Ends]

sp.07. Select the levels of analysis that apply to your measure.

Check ONLY the levels of analysis for which the measure is SPECIFIED and TESTED.
Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.
Please do not select:

- *Clinician: Clinician*
- *Population: Population*

[Response Begins]
Facility
[Response Ends]

sp.08. Indicate the care settings that apply to your measure.

Check ONLY the settings for which the measure is SPECIFIED and TESTED.
[Response Begins]
Inpatient/Hospital
[Response Ends]

sp.09. 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. If no URL is available, indicate "none available".

[Response Begins]
The specifications will be posted in the near future at https://www.jointcommission.org/measurement/specification-manuals/electronic-clinical-quality-measures/
[Response Ends]

sp.10. Indicate whether Health Quality Measure Format (HQMF) specifications are attached.

Attach the zipped output from the eCQM 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).
[Response Begins]
HQMF specifications are attached.
[Response Ends]

Attachment: 3687e_PC07_eCQMFlow2022.pdf
Attachment: 3687e_SOC-v0-0-138-QDM-5-6.zip

sp.11. Attach the data dictionary, code table, or value sets (and risk model codes and coefficients when applicable). Excel formats (.xlsx or .csv) are preferred.

Attach an excel or csv file; if this poses an issue, contact staff. Provide descriptors for any codes. Use one file with multiple worksheets, if needed.

[Response Begins]
Available in attached Excel or csv file
[Response Ends]

Attachment: 3687e_ePC07 eCQM Value Sets.xlsx

For the question below: state the outcome being measured. Calculation of the risk-adjusted outcome should be described in sp.22.

sp.12. State the numerator.

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.

[Response Begins]
Inpatient hospitalizations for patients with severe obstetric complications including the following:

- Severe maternal morbidity diagnoses (see list below)
- Severe maternal morbidity procedures (see list below)
- Discharge disposition = expired

Severe Maternal Morbidity Diagnoses:

- Cardiac
    - Acute heart failure
    - Acute myocardial infarction
    - Aortic aneurysm
    - Cardiac arrest/ventricular fibrillation
    - Heart failure/arrest during procedure or surgery
- Hemorrhage
    - Disseminated intravascular coagulation
    - Shock
- Renal
    - Acute renal failure
- Respiratory
    - Adult respiratory distress syndrome
    - Pulmonary edema
    - Sepsis
- Other OB
    - Air and thrombotic embolism
    - Amniotic fluid embolism
    - Eclampsia
    - Severe anesthesia complications
- Other Medical
- Puerperal cerebrovascular disorder
- Sickle cell disease with crisis

Severe Maternal Morbidity Procedures:

- Blood transfusion
- Conversion of cardiac rhythm
- Hysterectomy
- Temporary tracheostomy
- Ventilation

For further details on changes made to the numerator specifications during pilot testing, please see Changes Made to ePC07 Specifications During Pilot Testing in additional attachments.

[Response Ends]

For the question below: describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in sp.22.

sp.13. Provide details needed to calculate the numerator.

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 sp.11.

[Response Begins]
1. The QDM datatype of Encounter Performed, Diagnosis evaluates the Severe Maternal Morbidity Diagnoses value set (2.16.840.1.113762.1.4.1029.255) to see if a code is present on the encounter. If so, the Encounter, Performed, PresentOnAdmission Indicator datatype evaluates the Present on Admission = No or Unable to Determine value set (2.16.840.1.113762.1.4.1029.370) and the numerator will be met if the code has a POA code of "No" or "Unable to Determine".
2. The QDM datatype of Procedure, Performed evaluates the Severe Maternal Morbidity Procedures value set (2.16.840.1.113762.1.4.1029.256) and the Blood Transfusion value set (2.16.840.1.113762.1.4.1029.213) to see if a code is present with a corresponding procedure date anytime during the hospitalization encounter. The Blood Transfusion value set is kept separate from the other procedures so that the rates can be stratified with and without blood transfusion.
3. The QDM datatype of Encounter, Performed, Discharge Disposition evaluates the Patient Expired value set (2.16.840.1.113883.3.117.1.7.1.309) to determine if the patient expired during the encounter.
If any one of the 3 conditions above are met, the patient will be in the numerator. 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/ A list of value sets for the measure is attached in the Excel workbook provided for question sp.11.
For further details on changes made to the numerator specifications during pilot testing, please see Changes Made to ePC07 Specifications During Pilot Testing in additional attachments.

[Response Ends]

For the question below: state the target population for the outcome. Calculation of the risk-adjusted outcome should be described in sp.22.

sp.14. State the denominator.

Brief, narrative description of the target population being measured.

[Response Begins]
Initial Patient Population: Inpatient hospitalizations for patients age >= 8 years and < 65 admitted to the hospital for inpatient acute care who undergo a delivery procedure with a discharge date that ends during the measurement period
Denominator: Inpatient hospitalizations for patients delivering stillborn or live birth with >= 20 weeks, 0 days gestation completed

[Response Ends]

For the question below: describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in sp.22.

sp.15. Provide details needed to calculate the denominator.

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 sp.11.

[Response Begins]
For patients meeting the initial patient population:

1. The logic determines calculated gestational age (CGA) as follows:
    a. For the Estimated Due Date (EDD), the QDM datatype Assessment, Performed: Delivery date Estimated using Delivery date Estimated LOINC Direct Reference Code 11778-8 is used. To assure the most up to date EDD is used the logic looks for the last EDD 42 weeks or less before or on delivery.
    b. For the Date of Delivery, the QDM datatype Assessment, Performed: Date and time of obstetric delivery using Date and time of obstetric delivery LOINC Direct Reference Code 93857-1 is used. To assure the most accurate date/time of delivery the logic looks for the last assessment of date/time of delivery during the encounter. To account for deliveries that may occur outside of the inpatient encounter, the logic looks at the expanded encounter including any Emergency Department, Observation or OB Triage visits within one hour of the inpatient admission.
    c. The logic includes a function which calculates the gestational age. This function reflects the ACOG (American College of Obstetrics and Gynecology) ReVITALize Guidelines for Calculating Gestational Age (CGA):

        Gestational Age = (280-(EDD minus Reference Date))/7

Reference Date is the date on which you are trying to determine gestational age. For purposes of this eCQM, Reference Date would be the Date of Delivery.

2. If the necessary elements are not available to calculate CGA, CGA will be null. Then the estimated gestational age, which is derived from the QDM datatype Assessment, Performed: Estimated Gestational Age at Delivery using SNOMEDCT Value Set (2.16.840.1.113762.1.4.1045.26) is used.
3. Gestational age >= 20 weeks, 0 days will meet the logic.
4. Lastly, the QDM datatype of Procedure, Performed evaluates Procedure, Performed: Delivery Procedures (2.16.840.1.113762.1.4.1045.59) to determine if a delivery code is present. The delivery procedure codes do not distinguish live from stillborn deliveries.

[Response Ends]

sp.16. Describe the denominator exclusions.

Brief narrative description of exclusions from the target population.

[Response Begins]
Patients with confirmed diagnosis of COVID with COVID-related respiratory condition or patients with confirmed diagnosis of COVID with COVID-related respiratory procedure.
For further details on changes made to the denominator exclusion specifications during pilot testing please see Changes Made to ePC07 Specifications During Pilot Testing in additional attachments.

[Response Ends]

sp.17. Provide details needed to calculate the denominator exclusions.

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 sp.11.

[Response Begins]
A denominator exclusion for COVID plus respiratory conditions was added post pilot due to the growing evidence of perinatal complications in women who have COVID infection with respiratory conditions.
Patients with confirmed diagnosis of COVID with COVID-related respiratory condition or patients with confirmed diagnosis of COVID with COVID-related respiratory procedure are excluded.
1. The QDM datatype of Encounter Performed, Diagnosis evaluates the COVID 19 Confirmed value set (2.16.840.1.113762.1.4.1029.373) to see if a code is present on the encounter.
AND
2. The QDM datatype of Encounter Performed, Diagnosis evaluates the COVID 19 Related Respiratory Conditions value set (2.16.840.1.113762.1.4.1029.376) to see if a code is present on the encounter OR the QDM datatype of Procedure Performed evaluates COVID 19 Related Respiratory Procedures (2.16.840.1.113762.1.4.1029.379) and that the procedure starts during the encounter.
For further details on changes made to the denominator exclusion specifications during pilot testing please see Changes Made to ePC07 Specifications During Pilot Testing in additional attachments.

[Response Ends]

sp.18. Provide all information required to stratify the measure results, if necessary.

Include 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 in the Data Dictionary field.

[Response Begins]
A subset of the numerator population will be reported in Stratification as Stratum 1: Nontransfusion only severe obstetric complications (excluding cases where transfusion was the only severe obstetric complication)
Calculation:
(Risk-standardized number of encounters with nontransfusion only severe obstetric complications (excluding cases where transfusion was the only severe obstetric complication) / Number of encounters in Denominator) * 10,000
The logic includes a definition entitled: "Delivery Encounter Greater Than Or Equal To 20 Weeks Gestation Completed With Severe Obstetric Complications (Excluding Blood Transfusions)". This definition unions the following 2 definitions:

- "Delivery Encounter Greater Than Or Equal To 20 Weeks Gestation Completed With Severe Obstetric Complications Diagnosis or Procedure (Excluding Blood Transfusion)"
- Union "Delivery Encounter Greater Than Or Equal To 20 Weeks Gestation Completed With Expiration"

The first definition includes patients with a Severe Obstetric Complication Diagnosis or a procedure indicative of severe obstetric complication (other than blood transfusion) as described in the numerator. Cases with blood transfusions are not excluded from this definition if they have another SOC. Thereby, patients who only had a SOC of blood transfusion would not qualify for Stratum 1.

[Response Ends]

sp.19. Select the risk adjustment type.

Select type. Provide specifications for risk stratification and/or risk models in the Scientific Acceptability section.
[Response Begins]
Statistical risk model
[Response Ends]

sp.20. Select the most relevant type of score.

Attachment: If available, please provide a sample report.
[Response Begins]
Rate/proportion
[Response Ends]

sp.21. Select the appropriate interpretation of the measure score.

Classifies interpretation of score according to whether better quality or resource use is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score
[Response Begins]
Better quality = Lower score
[Response Ends]

sp.22. Diagram or describe the calculation of the measure score as an ordered sequence of steps.

Identify the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period of data, aggregating data; risk adjustment; etc.

[Response Begins]
Please see the attached HQMF specifications for the complete measure logic. Additionally, a flow diagram of the denominator, denominator exclusions, and numerator logic is attached to the NQF submission form as a supplemental document in response to question sp.10.
[Response Ends]

sp.25. If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.

[Response Begins]
No sampling.
[Response Ends]

sp.28. Select only the data sources for which the measure is specified.

[Response Begins]
Electronic Health Data
Electronic Health Records
[Response Ends]

sp.29. Identify the specific data source or data collection instrument.

For example, provide the name of the database, clinical registry, collection instrument, etc., and describe how data are collected.

[Response Begins]
Not applicable.
[Response Ends]

sp.30. Provide the data collection instrument.

[Response Begins]
No data collection instrument provided
[Response Ends]

Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate fields in the Scientific Acceptability sections of the Measure Submission Form.

- 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.
- All required sections must be completed.
- For composites with outcome and resource use measures, Questions 2b.23-2b.37 (Risk Adjustment) also must be completed.
- If specified for multiple data sources/sets of specifications (e.g., claims and EHRs), Questions 2b.11-2b.13 also must be completed.
- An appendix for supplemental materials may be submitted (see Question 1 in the Additional section), but there is no guarantee it will be reviewed.
- Contact NQF staff with any questions. Check for resources at the 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 the 2021 Measure Evaluation Criteria and Guidance.

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.
2a. Reliability testing demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For instrument-based measures (including PRO-PMs) and composite performance measures, reliability should be demonstrated for the computed performance score.
2b1. Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For 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;
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).
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 non-responders) and how the specified handling of missing data minimizes bias.
2c. For composite performance measures, empirical analyses support the composite construction approach and demonstrate that:
2c1. the component measures fit the quality construct and add value to the overall composite while achieving the related objective of parsimony to the extent possible; and

2c2. the aggregation and weighting rules are consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible.
(if not conducted or results not adequate, justification must be submitted and accepted)

Definitions
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).
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.
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.
Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.
Risk factors that influence outcomes should not be specified as exclusions.
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.

Please separate added or updated information from the most recent measure evaluation within each question response in the Importance to Scientific Acceptability sections. For example:
2021 Submission:
Updated testing information here.
2018 Submission:
Testing from the previous submission here.

2a.01. Select only the data sources for which the measure is tested.

[Response Begins]
Electronic Health Data
Electronic Health Records
[Response Ends]

2a.02. 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).

[Response Begins]
Not applicable.
[Response Ends]

2a.03. Provide the dates of the data used in testing.

Use the following format: "MM-DD-YYYY - MM-DD-YYYY"

[Response Begins]
01-01-2020 – 12-31-2020
[Response Ends]

2a.04. Select the levels of analysis for which the measure is tested.

Testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan.
Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.
Please do not select:

- *Clinician: Clinician*
- *Population: Population*

[Response Begins]
Facility
[Response Ends]

2a.05. List the measured entities 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.

[Response Begins]
enFor reference, each health system will be referred to as a 'pilot site' and 'hospital' will refer to the individual hospitals within the health system. A total of 10 pilot sites, consisting of 28 hospitals were included in some phase of pilot testing. For feasibility testing, 9 pilot sites with a total of 27 hospitals were included for analysis. After feasibility testing, 1 pilot site representing 2 hospitals withdrew from the project and one additional hospital was added. Therefore, data was collected from 9 pilot sites representing 26 hospitals. Reliability and validity testing was completed on 6 pilot sites representing 15 hospitals. One standalone hospital submitted data late in the process and therefore could not be included in the risk model development. However, their data is included in other analyses. See Table 2a.05.01 below for a summary of pilot site participation by testing phase.

Table 2a.05.01 Pilot Site Participation by Testing Phase

Testing Phase	# Pilot Sites	# Hospitals
Overall	10	28
Feasibility Testing	9	27
Data Collection	9	26
Reliability and Validity	6	15
Risk Model Development	8	25

Table 2a.05.01 displays the number of pilot sites and individual hospitals participating in each phase of testing. Table 2a.05.02 and Table 2a.05.03 provide health care system specific characteristics for each of the pilot sites. Table 2a.05.03 indicates whether a pilot site was included in feasibility testing, data collection and reliability/validity testing .

Table 2a.05.02 Pilot Site Characteristics

Pilot Site ID	# of Hospitals	State	Ownership Type*	Geography* (Urban, Suburban, Rural)	# Beds*	# of Births*	Teaching Program in OB/GYN*
Pilot Site 1	10	NC, VA	Nongovt. (not-for-profit) - Other	Urban	1807 (range 36 - 740)	1 6334 + (range 473 - 5568)	No

Pilot Site ID	# of Hospitals	State	Ownership Type*	Geography* (Urban, Suburban, Rural)	# Beds*	# of Births*	Teaching Program in OB/GYN*
Pilot Site 2	1	RI	Nongovt. (not-for-profit) - Other	Urban	247	8823	No
Pilot Site 3	1	LA	Nongovt. (not-for-profit) - Other	Urban	228	8295	No
Pilot Site 4[a]	2	CA	Nongovt. (not-for-profit) - Church Operated	Urban	446	2921	No
Pilot Site 5	9	OH, MI	Nongovt. (not-for-profit) – Other, Govt. (non federal) - County	6 Urban 3 Rural	1653 (range 35 - 595)	9283 + (range 165 – 3'Zz596)	No
Pilot Site 6	1	NJ	Nongovt. (not-for-profit) - Other	Urban	446	3319	No
Pilot Site 7	1	CA	Nongovt. (not-for-profit) - Other	Urban	541	4660	Yes
Pilot Site 8[b]	1	IL	Nongovt. (not-for-profit) - Other	Urban	650	2442	Yes
Pilot Site 9	1	MD	Nongovt. (not-for-profit) - Other	Urban	401	3854	No
Pilot Site 10[c]	1	PA	Nongovt. (not-for-profit) - Other	Urban	321	8796	Yes

*Source: American Hospital Association (AHA) DataQuery™ product, at the URL https://guide.prod.iam.aha.org/dataquery/reports, accessed March 16, 2021

a. Pilot Site 4 declined continued participation after Feasibility Testing

b. Data from Pilot Site 8 was collected but not available in time for Risk Model Development

c. Pilot Site 10 joined after Feasibility Testing

Table 2a.05.02 displays the characteristics of the entities measured. The information was retrieved from the American Hospital Association (AHA) DataQuery ™ product. For each pilot site, the table provides the number of hospitals, the state, ownership type, whether the site is located in an urban, suburban or rural setting, the number of beds, births and if the hospital has a teaching program in obstetrics and gynecology.

Table 2a.05.03 Pilot Site Characteristics Including Testing Phase Participation

Site ID	Obstetric unit care level*	NICU Level*	Clinical EHR Software	Included in Feasibility Testing	Included in Data Collection	Included in Reliability & Validity Testing
Pilot Site 1	*(Information not provided)*	Level 2 Level 3 Level 4	Epic	Yes	Yes	Yes
Pilot Site 2	Services all serious illnesses & abnormalities	Level 4	Cerner/ Siemens	Yes	Yes	Yes
Pilot Site 3	Services all serious illnesses & abnormalities	Level 3	Meditech	Yes	Yes	Yes

Site ID	Obstetric unit care level*	NICU Level*	Clinical EHR Software	Included in Feasibility Testing	Included in Data Collection	Included in Reliability & Validity Testing
Pilot Site 4[a]	Services uncomplicated maternity & newborn cases	Level 2 Level 3	Cerner	Yes	No	No
Pilot Site 5	*2 hospitals* = Services all serious illnesses & abnormalities *2 hospitals* = Services uncomplicated & most complicated cases *3 hospitals* = Services uncomplicated maternity & newborn cases *2 hospitals* = (Information not provided)	Level 3 (1 central NICU for all hospitals)	Epic	Yes	Yes	No
Pilot Site 6	Services all serious illnesses & abnormalities	Level 3	Meditech	Yes	Yes	Yes
Pilot Site 7	Services uncomplicated & most complicated cases	Level 3	Epic	Yes	Yes	Yes
Pilot Site 8[b]	Services all serious illnesses & abnormalities	Level 3	Epic	Yes	No	No
Pilot Site 9	Services all serious illnesses & abnormalities	Level 3	Epic	Yes	Yes	Yes
Pilot Site 10[c]	Services all serious illnesses & abnormalities	Level 3	Cerner	No	Yes	No

* Source: American Hospital Association (AHA) DataQuery™ product, at the URL https://guide.prod.iam.aha.org/dataquery/reports, accessed March 16, 2021

a. Pilot Site 4 declined continued participation after Feasibility Testing

b. Data from Pilot Site 8 was collected but not available in time for Risk Model Development

c. Pilot Site 10 joined after Feasibility Testing

Table 2a.05.03 displays the entity characteristics for obstetric level, NICU level, EHR software, and the phase of testing the hospital participated in. 3 EHR software systems, Meditech, EPIC, and Cerner, were tested.

[Response Ends]

2a.06. Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis), separated by level of analysis and data source; if a sample was used, describe how patients were selected for inclusion in the sample.

If there is a minimum case count used for testing, that minimum must be reflected in the specifications.

[Response Begins]

Table 2a.06.01 Patient Characteristics for 8 Pilot Sites Participating in Data Collection Phase

Category	Site #1	Site #2	Site #3	Site #5	Site #6	Site #7	Site #9	Site #10	Across Sites
*	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
Number of encounters	18,070	7,196	7,955	6,139	3,359	4,369	3,918	9,178	60,184
Number of unique patients	18,070	7,196	7,949	6,139	3,359	4,367	3,918	9,173	60,170

Category	Site #1	Site #2	Site #3	Site #5	Site #6	Site #7	Site #9	Site #10	Across Sites
Average Maternal Age in Years [Mean (STD)]	30 (6.0)	31 (6.0)	29 (6.0)	29 (6.0)	33 (5.0)	32 (5.0)	32 (5.0)	31 (5.0)	30 (6.0)
Maternal Age in Years	*	*	*	*	*	*	*	*	*
<18	111 (0.6)	39 (0.5)	78 (1.0)	51 (0.8)	1 (0.0)	2 (0.0)	10 (0.3)	52 (0.6)	344 (0.6)
18-<25	3158 (17.5)	1130 (15.7)	1822 (22.9)	1530 (24.9)	145 (4.3)	391 (8.9)	356 (9.1)	1255 (13.7)	9787 (16.3)
25-<30	4917 (27.2)	1791 (24.9)	2416 (30.4)	1885 (30.7)	490 (14.6)	959 (22.0)	860 (21.9)	2194 (23.9)	15512 (25.8)
30-<35	5908 (32.7)	2413 (33.5)	2223 (27.9)	1708 (27.8)	1417 (42.2)	1622 (37.1)	1542 (39.4)	3404 (37.1)	20237 (33.6)
35-<40	3161 (17.5)	1458 (20.3)	1177 (14.8)	800 (13.0)	1007 (30.0)	1118 (25.6)	914 (23.3)	1864 (20.3)	11499 (19.1)
40-<45	749 (4.1)	341 (4.7)	223 (2.8)	153 (2.5)	277 (8.2)	263 (6.0)	215 (5.5)	387 (4.2)	2608 (4.3)
45-<50	60 (0.3)	21 (0.3)	15 (0.2)	12 (0.2)	19 (0.6)	13 (0.3)	19 (0.5)	18 (0.2)	177 (0.3)
>=50	6 (0)	3 (0.0)	0 (0.0)	0 (0.0)	3 (0.1)	1 (0.0)	2 (0.1)	4 (0.0)	19 (0.0)
Race/Ethnicity	*	*	*	*	*	*	*	*	*
Hispanic	2468 (13.7)	2110 (29.3)	734 (9.2)	485 (7.9)	497 (14.8)	1739 (39.8)	163 (4.2)	235 (2.6)	8431 (14.0)
Non-Hispanic, African American	4084 (22.6)	606 (8.4)	2971 (37.3)	952 (15.5)	89 (2.6)	254 (5.8)	1307 (33.4)	1590 (17.3)	11853 (19.7)
Non-Hispanic, Asian/Pacific Islander	743 (4.1)	117 (1.6)	157 (2.0)	66 (1.1)	364 (10.8)	703 (16.1)	250 (6.4)	532 (5.8)	2932 (4.9)
Non-Hispanic, White	9322 (51.6)	3658 (50.8)	3940 (49.5)	4507 (73.4)	2307 (68.7)	1648 (37.7)	2077 (53.0)	5912 (64.4)	33371 (55.4)
Non-Hispanic, Other	651 (3.6)	633 (8.8)	135 (1.7)	58 (0.9)	35 (1.0)	17 (0.4)	112 (2.9)	40 (0.4)	1681 (2.8)
Declined/Unknown	802 (4.4)	72 (1.0)	18 (0.2)	71 (1.2)	67 (2.0)	8 (0.2)	9 (0.2)	869 (9.5)	1916 (3.2)
Primary Payer	*	*	*	*	*	*	*	*	*
Medicare	50 (0.3)	12 (0.2)	27 (0.3)	36 (0.6)	7 (0.2)	1 (0.0)	6 (0.2)	84 (0.9)	223 (0.4)
Medicaid	5857 (32.4)	305 (4.2)	3790 (47.6)	2600 (42.4)	97 (2.9)	408 (9.3)	10 (0.3)	3154 (34.4)	16221 (27.0)
Private Insurance	11170 (61.8)	6863 (95.4)	4119 (51.8)	3482 (56.7)	3230 (96.2)	3869 (88.6)	3894 (99.4)	4439 (48.4)	41066 (68.2)
Self-pay or Uninsured	0 (0.0)	15 (0.2)	19 (0.2)	21 (0.3)	15 (0.4)	0 (0.0)	8 (0.2)	71 (0.8)	149 (0.2)
Other	993 (5.5)	0 (0.0)	0 (0.0)	0 (0.0)	10 (0.3)	86 (2.0)	0 (0.0)	1429 (15.6)	2518 (4.2)
Unknown	0 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	5 (0.1)	0 (0.0)	1 (0.0)	7 (0.0)

* Indicates the table cell left intentionally blank

Table 2a.06.01 displays the patient characteristics for 8 pilot sites used in data collection. Each site has the number of encounters, number of unique patients, and the average maternal age in years. The table also displays a breakdown of maternal age in years, race and ethnicity and payer categories.

[Response Ends]

2a.07. 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.

[Response Begins]
As described in 2a.05, a total of 10 pilot sites, consisting of 28 hospitals were included in some phase of pilot testing. For feasibility testing, 9 pilot sites with a total of 27 hospitals were included for analysis. After feasibility testing, 1 pilot site representing 2 hospitals withdrew from the project and one additional hospital was added. Therefore, data was collected from 9 pilot sites representing 26 hospitals. Reliability and validity testing was completed on 6 pilot sites representing 15 hospitals. One standalone hospital submitted data late in the process and therefore could not be included in the risk model development. However, their data is included in other analyses. See above Table 2a.05.01 Pilot Site Participation by Testing Phase and Table 2a.05.03 Pilot Site Characteristics Including Testing Phase Participation for a summary of pilot site participation by testing phase.
[Response Ends]

2a.08. List 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.

[Response Begins]
Economic/housing instability was included in the risk model. Race/ethnicity was examined as a stratification variable rather than risk variables. It was determined that illumination of outcome disparities by race/ethnicity, rather than adjustment of outcomes by race/ethnicity, would best inform stakeholders and patients and be most impactful in incentivizing improvements in quality of maternal care. Analysis on how best to report stratification for race/ethnicity is ongoing.
[Response Ends]

Note: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a.07 check patient or encounter-level data; in 2a.08 enter "see validity testing section of data elements"; and enter "N/A" for 2a.09 and 2a.10.

2a.09. Select the level of reliability testing conducted.

Choose one or both levels.
[Response Begins]
Patient or Encounter-Level (e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements)
Accountable Entity Level (e.g., signal-to-noise analysis)
[Response Ends]

2a.10. For each level of reliability testing 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.

[Response Begins]
Accountable Entity Level ("Measure Score") Reliability: Measure scores were calculated for 8 pilot sites used for risk model development, and for the 25 individual hospitals within those 8 pilot sites. During measure testing, we assessed measure score reliability, which is the degree to which repeated measurements of the same entity agree with each other. We estimated the measure score reliability using a signal-to-noise ratio to assess the values according to conventional standards (Landis & Koch, 1977). We assessed signal-to-noise reliability that describes how well the measure can distinguish the performance of one hospital from another. The signal is the proportion of the variability in measured performance that can be explained by real differences in performance. Scores can range from zero to one. A reliability of zero implies that all the variability in a measure is attributable to measurement error. A reliability of one implies that all the variability is attributable to real difference in performance (Yu et al., 2013; Adams et al., 2010).

For measure score reliability testing of measure scores for the 25 individual hospitals, testing was conducted at several volume thresholds, including: no required minimum number of delivery encounters for the year, at least 25 delivery encounters for the year, and 200 delivery encounters for the year. None of the testing hospitals had fewer than 25 delivery encounters, so results at the no minimum delivery encounter volume threshold and 25-delivery encounter threshold are the same.

Landis, J.R. & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*.159-174.
Yu, H., Mehrotra, A., & Adams, J. (2013). Reliability of utilization measures for primary care physician profiling. Paper presented at: Healthcare2013.
Adams, J.L., Mehrotra, A., Thomas, J.W., & McGlynn, E.A. (2010). Physician cost profiling—reliability and risk of misclassification. *New England Journal of Medicine*, 362(11), 1014-1021.

[Response Ends]

2a.11. For each level of reliability testing checked above, what were the statistical results from reliability testing?

For example, provide the percent agreement and kappa for the critical data elements, or distribution of reliability statistics from a signal-to-noise analysis. For score-level reliability testing, when using a signal-to-noise analysis, more than just one overall statistic should be reported (i.e., to demonstrate variation in reliability across providers). If a particular method yields only one statistic, this should be explained. In addition, reporting of results stratified by sample size is preferred (pg. 18, NQF Measure Evaluation Criteria).

[Response Begins]

Accountable Entity Level ("Measure Score") Reliability: Results at the health site level (N=8 pilot sites), presented in Table 2a.11.01, indicate that this reliability analysis yielded a median reliability score of 0.991 (range: 0.982 – 0.997) for any severe obstetric complication and 0.955 (range: 0.916 – 0.983) for severe obstetric complications excluding blood transfusion-only cases. Each pilot site had at least 25 delivery encounters.

Signal-to-noise reliability was calculated for the 25 individual hospitals at several volume thresholds; results are provided for hospitals with at least 25 delivery encounters in the year (all hospitals included in testing) and for hospitals with at least 200 delivery encounters in the year (23 of the 25 hospitals included in testing). For hospitals with at least 25 delivery encounters (see Table 2a.11.02), the median reliability score was 0.959 (0.802-0.996) for any severe obstetric complication outcome and 0.684 (0.273-0.961) for severe obstetric complications excluding blood transfusion-only cases. The signal-to-noise reliability is higher when included hospitals had at least 200 delivery encounters for the year, rather than 25 delivery encounters, particularly for the second outcome (severe complications excluding blood transfusion-only cases: the median reliability score was 0.978 (0.867-0.996) for any severe obstetric complication outcome and 0.804 (0.377-0.961) for severe obstetric complications excluding blood transfusion-only cases.

Table 2a.11.01. Signal-to-Noise-Reliability, Measure Scores, Site Level

*	# Pilot Sites	Median	Mean	Minimum	Maximum	Interquartile Range Q1	Interquartile Range Q3
Any Severe Obstetric Complication(s)	8	0.991	0.99 (0.005)	0.982	0.997	0.985	0.993
Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases	8	0.955	0.95 (0.023)	0.916	0.983	0.929	0.966

* Indicates the table cell left intentionally blank

Table 2a.11.01 displays the signal-to-noise reliability, measure scores for the site level. The mean is 0.99 for any severe obstetric complications among the 8 pilot sites and 0.95 for the severe obstetric complications excluding blood transfusion only cases.

Table 2a.11.02. Signal-to-Noise-Reliability, Measure Scores, Hospital Level

*	Volume Threshold (Number of Delivery Encounters per Hospital per year)	# Pilot Hospitals	Median	Mean	Minimum (SD)	Maximum	Interquartile Range Q1	Interquartile Range Q3
Any Severe Obstetric Complication(s)	**>25**	25	0.959	0.946 (0.056)	0.802	0.996	0.904	0.991
Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases	**>25**	25	0.684	0.694 (0.229)	0.273	0.961	0.466	0.913
Any Severe Obstetric Complication(s)	**>200**	23	0.978	0.958 (0.040)	0.867	0.996	0.935	0.991
Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases	**>200**	23	0.804	0.729 (0.202)	0.377	0.961	0.573	0.914

* Indicates the table cell left intentionally blank

Table 2a.11.02 displays the signal-to-noise reliability measure scores at the hospital level using two thresholds, >25 and >200. The mean measure scores for any severe obstetric complication are 0.959 for a threshold of >25 and 0.978 for a threshold of >200. The mean for severe obstetric complications excluding blood transfusion-only cases are 0.684 for a threshold of >25 and 0.804 for a threshold >200.

[Response Ends]

2a.12. Interpret the results, in terms of how they demonstrate reliability.

(In other words, what do the results mean and what are the norms for the test conducted?)

[Response Begins]
Accountable Entity Level ("Measure Score") Reliability: The signal-to-noise reliability results at the health site level show very high reliability for both outcomes. Signal-to-noise reliability results at the hospital level indicate very high reliability for the outcome measuring any severe obstetric complication, and a more moderate reliability for the outcome measuring any severe obstetric complications excluding blood transfusion. Setting a minimum threshold of at least 200 delivery encounters per hospital increases reliability, particularly for the severe obstetric complications excluding transfusion-only cases, which impacts fewer patients and represents a rarer outcome.
[Response Ends]

2b.01. Select the level of validity testing that was conducted.

[Response Begins]
Patient or Encounter-Level (data element validity must address ALL critical data elements)
Accountable Entity Level (e.g. hospitals, clinicians)

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)
[Response Ends]

2b.02. 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.

[Response Begins]
Validity testing was completed for 15 individual hospitals at 6 pilot sites. This includes 1 system of 10 hospitals and 5 individual hospitals. We reviewed 3-4 charts for each hospital in the system and 30-36 charts at each of the individual hospitals. The review included three different EHR vendors, 3 sites use Epic, 2 use Meditech and 1 site uses Cerner. Due to COVID-19, onsite validity testing visits were transitioned to a virtual visit approach. Validity testing was conducted April through July of 2021 by The Joint Commission staff with the support of a hospital site abstractor. The purpose of the visits was to assess data element validity through clinical adjudication; elicit feedback from pilot site staff as to the importance, feasibility, and usability of the measure data elements, as well as determine if measure specifications were sufficiently clear and detailed to promote comparability of measure findings across hospitals.
A. Re-abstraction/Clinical Adjudication
A statistically representative sample of the electronically submitted inpatient encounters was selected for re-abstraction. During the virtual visits, site staff shared their screen, navigated through the electronic health records of the sampled patients while The Joint Commission staff manually re-abstracted each data element. To determine validity, re-abstraction findings were compared with the original electronic data submission and any disagreements were adjudicated with reasons for discrepancies noted.
B. Analysis
Validity testing methodology was done as outlined below:

- All clinical data elements and all editable demographic elements are scored.
- All measure data are re-abstracted with original data having been blinded so that the re-abstraction is not biased.
- Re-abstracted data are compared with original data for each data element. Ideally, data element agreement rates should exceed 80%.
- Overall performance measure outcome rates were calculated on all cases submitted by each pilot site. Next, performance measure outcome rates were calculated on the adjudicated data for the sampled cases. The performance measure outcome rates were compared, and agreement rates were corrected for chance variation with the kappa statistic. Ideally, a kappa score greater than .60 should be achieved.

When assessing agreement, we used the following kappa score ranges as guidance:

- < 0: Less than chance agreement
- 0.01–0.20: Slight agreement
- 0.21– 0.40: Fair agreement
- 0.41–0.60: Moderate agreement
- 0.61–0.80: Substantial agreement
- 0.81–0.99: Almost perfect agreement

To assess face validity, a Qualtrics survey was produced and distributed to the members of the Technical Expert Panel (TEP) for their completion. Members were asked to rate the following statements:

1. The severe obstetric morbidity and mortality captured by the Severe Obstetric Complications eCQM is an important health outcome to measure because it is an area with room for improvement.
2. The Severe Obstetric Complications eCQM will produce reliable and valid hospital measurement of severe obstetric morbidity and mortality rates across hospitals.
3. The Severe Obstetric Complications eCQM is feasible to implement because required data are routinely collected as part of clinical care and are extractable from electronic health records.
4. Hospitals can use the Severe Obstetric Complications eCQM performance results for performance improvement.
5. The risk standardized rate of severe obstetric morbidity and mortality events obtained from the Severe Obstetric Complications eCQM as specified is a critical component (that is, necessary but not all-inclusive) of defining and comparing quality of obstetric care between hospitals.

A Likert scale was used with the 6 possible responses ranging from Strongly Agree to Strongly Disagree. Each statement included an opportunity for the respondent to provide additional rationale if a disagree response was added.
In addition, the Patient Working Group was sent a Qualtrics survey to assess face validity with two of the five questions listed above (Questions 1 and 5), with a Likert scale was used with the 6 possible responses ranging from Strongly Agree to Strongly Disagree.

[Response Ends]

2b.03. Provide the statistical results from validity testing.

Examples may include correlations or t-test results.

[Response Begins]
Accountable Entity Level "Measure Score Validity": Measure score validity testing was completed in the same pilot sites. Table 2b.03.01 displays the PPV (agreement rate) for the numerator among delivery encounters clinically adjudicated in validity testing. The PPV rate was 100% at Pilot Sites 1, 2, 3, 6, and 7, and 70% at Pilot Site 9, with an overall PPV of 94.74%.

Table 2b.03.01. Agreement Statistics for Measure Numerator between EHR Extraction and Manual Chart Abstraction (PPV) (Validity Testing, 6 Pilot Sites)

Pilot Sites	# Of Numerator Events Verified by Clinical Adjudication	# Of Numerator Events from EHR	Positive Predictive Value (PPV)
*	*	*	*
*	*	*	*
Pilot Site 1	20	20	100%
Pilot Site 2	16	16	100%
Pilot Site 3	20	20	100%
Pilot Site 6	20	20	100%
Pilot Site 7	18	18	100%
Pilot Site 9	14	20	70.00%
Across 6 Pilot Sites	108	114	94.74%

* Indicates the table cell left intentionally blank

Table 2b.03.01 displays the PPV (agreement rate) for the numerator among delivery encounters clinically adjudicated in validity testing. The PPV rate was 100% at Pilot Sites 1, 2, 3, 6, and 7, and 70% at Pilot Site 9, with an overall PPV of 94.74%.
Table 2b.03.02 displays the sensitivity, specificity, and negative predictive value (NPV). Specificity and sensitivity are high. Sensitivity is 100% in all reliability pilot sites and specificity is 100% in pilot sites 1, 2, 3, 6, and 7 and 62.5% in pilot site 9. This means that the probability of the EHR data detecting a true severe obstetric complication during a delivery hospitalization based on the abstracted data ('gold standard') is 100% (sensitivity). The probability of the EHR data accurately identifying that no severe obstetric complication occurred during a delivery hospitalization based on abstracted data ranged from 62.5% to 100% and was 90.48% across pilot sites (specificity). NPV was 100% in all pilot sites, indicating the EHR data indicated a severe obstetric complication did not occur, and 100% of the time the chart abstraction confirmed a severe obstetric complication did not occur.

Table 2b.03.02. Measure Score Validity Statistics for Sample Between EHR Extraction and Manual Chart Abstraction (Sensitivity, Specificity, NPV)

Pilot Sites	Sensitivity	Specificity	Negative Predictive Value (NPV)
*	*	*	*
*	*	*	*
Pilot Site 1	100%	100%	100%

Pilot Sites	Sensitivity	Specificity	Negative Predictive Value (NPV)
Pilot Site 2	100%	100%	100%
Pilot Site 3	100%	100%	100%
Pilot Site 6	100%	100%	100%
Pilot Site 7	100%	100%	100%
Pilot Site 9	100%	62.50%	100%
Across 6 Pilot Sites	100%	90.48%	100%

* Indicates the table cell left intentionally blank

Table 2b.03.02 displays the measure score validity statistics for pilot site samples between the EHR extraction and the manual chart abstraction. Across 6 pilot sites, the sensitivity was 100%, the specificity was 90.48% and the NPV was 100%.

Overall, the study revealed ePC-07 to have an excellent measure outcome agreement rate of 91.2% with a kappa score of 0.881 indicating almost perfect agreement. (See Table 2b.03.03 Measure Outcome Agreement Rates.)

Table 2b.03.03 Measure Outcome Agreement Rates

Pilot Sites	N	Agreement Rate	kappa
Pilot Site 1	36	97.2%	0.963
Pilot Site 2	31	83.9%	0.786
Pilot Site 3	35	94.3%	0.922
Pilot Site 6	36	97.2%	0.963
Pilot Site 7	30	96.7%	0.953
Pilot Site 9	36	77.8%	0.703
Total	204	91.2%	0.881

Table 2b.03.03 displays the measure outcome agreement rates for the pilot sites. The total agreement rate is 91.2% and the kappa is 0.881.

Overall, the data element agreement rate for all sites was excellent at a score of 90.4%. (See Table 2b.03.04 Data Element Agreement Rates.)

Table 2b.03.04 Data Element Agreement Rates

*	Site 1	Site 1	Site 1	Site 2	Site 2	Site 2	Site 3	Site 3	Site 3	Site 6	Site 6	Site 6	Site 7	Site 7	Site 7	Site 9	Site 9	Site 9	Total	Total	Total
Data Element Name	**Match**	**N**	**Rate**	**Match**	**N**	**Rate**	**Match**	**N**	**Rate**	**Match**	**N**	**Rate**	**Match**	**N**	**Rate**	**Match**	**N**	**Rate**	**Match**	**N**	**Rate**
Demographics	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
DOB	36	36	100%	31	31	100%	35	35	100%	36	36	100%	30	30	100%	36	36	100%	204	204	100%

*	Site 1	Site 1	Site 1	Site 2	Site 2	Site 2	Site 3	Site 3	Site 3	Site 6	Site 6	Site 6	Site 7	Site 7	Site 7	Site 9	Site 9	Site 9	Total	Total	Total
ONC Administrative Sex Code	36	36	100%	31	31	100%	35	35	100%	36	36	100%	30	30	100%	36	36	100%	204	204	100%
Race	36	36	100%	31	31	100%	35	35	100%	35	35	100%	30	30	100%	36	36	100%	203	203	100%
Ethnicity	36	36	100%	31	31	100%	35	35	100%	35	35	100%	30	30	100%	36	36	100%	203	203	100%
Payer	32	36	89%	31	31	100%	35	35	100%	36	36	100%	30	30	100%	36	36	100%	200	204	98%
Admission Source	34	36	94%	31	31	100%	35	35	100%	36	36	100%	30	30	100%	36	36	100%	202	204	99%
dischargeDisposition	35	36	97%	30	31	97%	35	35	100%	36	36	100%	30	30	100%	36	36	100%	202	204	99%
Encounter History	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Encounter, Performed: Encounter Inpatient	36	36	100%	31	31	100%	35	35	100%	36	36	100%	30	30	100%	36	36	100%	204	204	100%
Admission Date Time (Relevant Period Start Time)	36	36	100%	30	31	97%	35	35	100%	35	36	97%	1	30	3%	36	36	100%	173	204	85%
Discharge Date Time (Relevant Period End Time)	36	36	100%	31	31	100%	35	35	100%	36	36	100%	30	30	100%	36	36	100%	204	204	100%

*	Site 1	Site 1	Site 1	Site 2	Site 2	Site 2	Site 3	Site 3	Site 3	Site 6	Site 6	Site 6	Site 7	Site 7	Site 7	Site 9	Site 9	Site 9	Total	Total	Total
Encounter, Performed: Emergency Department Visit	6	6	100%	12	12	100%	16	16	100%	0	0	*	0	0	*	3	3	100%	37	37	100%
ED Start Date Time (relevant Period)	6	6	100%	12	12	100%	16	16	100%	0	0	*	0	0	*	3	3	100%	37	37	100%
ED End Date Time (relevant Period)	6	6	100%	12	12	100%	16	16	100%	0	0	*	0	0	*	3	3	100%	37	37	100%
Encounter, Performed: Preadmission Observation Undelivered Mother	0	0	*	0	7	0%	0	0	*	9	9	100%	0	29	0%	36	36	100%	45	81	56%
PreAdm Obs Start Date Time (relevant Period)	0	0	*	0	7	0%	0	0	*	9	9	100%	0	29	0%	36	36	100%	45	81	56%
PreAdm Obs End Date Time (relevant Period)	0	0	*	0	7	0%	0	0	*	9	9	100%	0	29	0%	36	36	100%	45	81	56%
Encounter, Performed: Observation Services	25	27	93%	1	1	100%	0	0	*	1	1	100%	0	0	*	36	36	100%	63	65	97%

*	Site 1	Site 1	Site 1	Site 2	Site 2	Site 2	Site 3	Site 3	Site 3	Site 6	Site 6	Site 6	Site 7	Site 7	Site 7	Site 9	Site 9	Site 9	Total	Total	Total
Obs Start Date Time (relevant Period)	25	27	93%	1	1	100%	0	0	*	1	1	100%	0	0	*	36	36	100%	63	65	97%
Obs End Date Time (relevant Period)	25	27	93%	1	1	100%	0	0	*	1	1	100%	0	0	*	36	36	100%	63	65	97%
facilityLocations: Intensive Care Unit Code	2	2	100%	0	0	*	5	5	100%	1	1	100%	0	0	*	0	0	*	8	8	100%
ICU Start Date Time	2	2	100%	0	0	*	5	5	100%	1	1	100%	0	0	*	0	0	*	8	8	100%
ICU End Date Time	2	2	100%	0	0	*	5	5	100%	1	1	100%	0	0	*	0	0	*	8	8	100%
Dx	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Diagnosis POA	245	391	63%	397	397	100%	312	312	100%	208	346	60%	319	319	100%	327	328	100%	1808	2093	86%
Diagnosis code	391	391	100%	397	397	100%	312	312	100%	346	346	100%	319	319	100%	328	328	100%	2093	2093	100%
Proc	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Procedure code & date	103	104	99%	140	142	99%	114	115	99%	103	103	100%	93	93	100%	78	79	99%	631	636	99%
Blood	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Blood Transfusion code	33	33	100%	27	148	18%	31	31	100%	31	31	100%	10	15	67%	19	20	95%	151	278	54%
Blood Transfusion start	33	33	100%	25	141	18%	31	31	100%	31	31	100%	10	15	67%	19	20	95%	149	271	55%
Blood Transfusion end	25	33	76%	24	138	17%	31	31	100%	30	30	100%	5	15	33%	19	20	95%	134	267	50%
Delivery Details	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*

*	Site 1	Site 1	Site 1	Site 2	Site 2	Site 2	Site 3	Site 3	Site 3	Site 6	Site 6	Site 6	Site 7	Site 7	Site 7	Site 9	Site 9	Site 9	Total	Total	Total
Relevant Date Time Assessment, Performed: Date and time of obstetric delivery	35	36	97%	30	31	97%	34	35	97%	36	36	100%	28	30	93%	36	36	100%	199	204	98%
Result: Date and time of obstetric delivery	35	36	97%	30	31	97%	34	35	97%	36	36	100%	28	30	93%	36	36	100%	199	204	98%
Relevant Date Time Assessment, Performed: Delivery date Estimated	34	36	94%	0	31	0%	34	35	97%	36	36	100%	28	30	93%	36	36	100%	168	204	82%
Result: Delivery date Estimated	34	36	94%	30	31	97%	35	35	100%	36	36	100%	30	30	100%	34	36	94%	199	204	98%
Relevant Date Time Assessment, Performed: Estimated Gestational Age at Delivery	35	36	97%	24	30	80%	34	35	97%	36	36	100%	28	30	93%	34	36	94%	191	203	94%
Result: Estimated Gestational Age at Delivery	36	36	100%	24	31	77%	35	35	100%	36	36	100%	30	30	100%	34	36	94%	195	204	96%

*	Site 1	Site 1	Site 1	Site 2	Site 2	Site 2	Site 3	Site 3	Site 3	Site 6	Site 6	Site 6	Site 7	Site 7	Site 7	Site 9	Site 9	Site 9	Total	Total	Total
Laboratory Results	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Creatinine Result Date Time	0	0	*	0	0	*	2	2	100%	0	0	*	1	1	100%	0	0	*	3	3	100%
Creatinine Result	0	0	*	0	0	*	2	2	100%	0	0	*	1	1	100%	0	0	*	3	3	100%
PaO2 Result Date Time	0	0	*	1	1	100%	0	0	*	0	0	*	0	0	*	2	10	20%	3	11	27%
PaO2 Result	0	0	*	1	1	100%	0	0	*	0	0	*	0	0	*	2	10	20%	3	11	27%
Platelet Result Date Time	9	9	100%	4	4	100%	3	3	100%	9	9	100%	8	9	89%	2	2	100%	35	36	97%
Platelet Result	9	9	100%	4	4	100%	3	3	100%	9	9	100%	8	9	89%	2	2	100%	35	36	97%
Hemoglobin Result Date Time	117	117	100%	98	99	99%	89	89	100%	108	109	99%	69	92	75%	50	50	100%	531	556	96%
Hemoglobin Result	117	117	100%	98	99	99%	89	89	100%	108	109	99%	71	92	77%	50	50	100%	533	556	96%
Hematocrit Result Date Time	117	117	100%	97	99	98%	93	93	100%	108	109	99%	70	92	76%	111	112	99%	596	622	96%
Hematocrit Result	117	117	100%	98	99	99%	93	93	100%	108	109	99%	70	92	76%	111	112	99%	597	622	96%
WBC Result Date Time	105	105	100%	97	99	98%	92	92	100%	108	109	99%	70	92	76%	49	49	100%	521	546	95%
WBC Result	105	105	100%	98	99	99%	92	92	100%	108	109	99%	70	92	76%	49	49	100%	522	546	96%

*	Site 1	Site 1	Site 1	Site 2	Site 2	Site 2	Site 3	Site 3	Site 3	Site 6	Site 6	Site 6	Site 7	Site 7	Site 7	Site 9	Site 9	Site 9	Total	Total	Total
Glucose Result Date Time	19	19	100%	16	32	50%	31	31	100%	27	28	96%	1	9	11%	16	28	57%	110	147	75%
Glucose Result	19	19	100%	16	32	50%	31	31	100%	27	28	96%	1	9	11%	16	28	57%	110	147	75%
Bicarbonate Result Date Time	0	11	0%	6	6	100%	27	27	100%	0	26	0%	5	6	83%	14	14	100%	52	90	58%
Bicarbonate Result	0	11	0%	6	6	100%	27	27	100%	0	26	0%	5	6	83%	14	14	100%	52	90	58%
Vital Signs	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Relevant Date Time Physical Exam, Performed: Oxygen saturation in Arterial blood by Pulse oximetry (%)	4	35	11%	19	27	70%	34	34	100%	34	34	100%	29	29	100%	31	31	100%	151	190	80%
Result: Oxygen saturation	34	35	97%	21	27	78%	34	34	100%	34	34	100%	29	29	100%	31	31	100%	183	190	96%
Relevant Date Time Physical Exam, Performed: Heart rate (BPM)	12	36	33%	19	31	61%	35	35	100%	31	35	89%	28	30	93%	36	36	100%	161	203	79%
Result: Heart rate	36	36	100%	23	31	74%	35	35	100%	31	35	89%	28	30	93%	36	36	100%	189	203	93%

*	Site 1	Site 1	Site 1	Site 2	Site 2	Site 2	Site 3	Site 3	Site 3	Site 6	Site 6	Site 6	Site 7	Site 7	Site 7	Site 9	Site 9	Site 9	Total	Total	Total
Relevant Date Time Physical Exam, Performed: Systolic blood pressure (mmHg)	12	36	33%	19	31	61%	35	35	100%	31	35	89%	28	30	93%	36	36	100%	161	203	79%
Result: Systolic blood pressure	36	36	100%	23	31	74%	35	35	100%	31	35	89%	28	30	93%	36	36	100%	189	203	93%
Relevant Date Time Physical Exam, Performed: Respiratory rate (breaths per minute)	10	36	28%	19	31	61%	35	35	100%	22	24	92%	29	30	97%	36	36	100%	151	192	79%
Result: Respiratory rate	35	36	97%	23	31	74%	35	35	100%	22	24	92%	29	30	97%	36	36	100%	180	192	94%
Relevant Date Time Physical Exam, Performed: Body temperature (degrees Fahrenheit or degrees Celsius)	7	36	19%	19	31	61%	35	35	100%	29	32	91%	29	30	97%	36	36	100%	155	200	78%
Result: Body temperature	36	36	100%	23	31	74%	35	35	100%	29	32	91%	29	30	97%	36	36	100%	188	200	94%
TOTALS	**2447**	**2780**	**88%**	**2343**	**2900**	**81%**	**2472**	**2477**	**100%**	**2369**	**2594**	**91%**	**1935**	**2243**	**86%**	**2423**	**2476**	**98%**	**13989**	**15470**	**90%**

*This cell intentionally left empty.

Table 2b.03.04 displays the data element agreement rates for each pilot site. The individual data element match rate is provided as well as the total match rate of all data elements for each pilot site.
The Technical Expert Panel Face Validity Results and The Patient Working Group Panel Face Validity Results can be found as attachments in the Additional section.
15 members of the TEP completed face validity surveys. 80% of TEP members strongly agree while 20% moderately agree that this is an important health outcome to measure because there is room for improvement. 87% strongly or moderately agree the eCQM will produce reliable and valid rates while the remaining 13% of respondents somewhat agree. Similarly, 87% strongly or moderately agree that hospitals can use the results for performance improvement, while the remaining 13% of respondents somewhat agree.
The majority of members (12) agreed while 3 members disagreed with statement 3 which assesses implementation feasibility. Those that disagreed indicated that implementation will require additional resources but support the transition to eCQMs. One respondent stated demographic data like race and ethnicity are not routinely collected in all hospitals. Feasibility testing revealed that race and ethnicity data elements are routinely collected; however, there is not standardization amongst hospitals.
The majority of members (12) agreed while 3 members disagreed with statement 5 which assesses if the rate is a critical component of defining and comparing quality of obstetric care between hospitals. Those that disagreed indicated that the metric needs to be tested and stratified before risk adjusting. Other comments indicated the need to assess individual case results to confirm that outcomes reflect quality of care.
Five members of the Patient Working Group (PWG) completed the face validity surveys with two of the five statements. All five Patient Working Group members strongly agreed with the first statement ("The severe obstetric morbidity and mortality captured by the Severe Obstetric Complications eCQM is an important health outcome to measure because it is an area with room for improvement"). The second statement ("The risk standardized rate of severe obstetric morbidity and mortality events obtained from the Severe Obstetric Complications eCQM as specified is a critical component (that is, necessary but not all-inclusive) of defining and comparing quality of obstetric care between hospitals") was rated by all respondents as strongly (n=3)/moderately agree (n=2). These results demonstrate that the PWG believes this is an important health outcome to measure because there is room for improvement and the rate is a critical component of defining and comparing quality of obstetric care between hospitals.

[Response Ends]

2b.04. Provide 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?)

[Response Begins]
Accountable Entity Level "Measure Score Validity":
Positive Predictive Value (PPV): In almost all delivery encounters with a numerator event adjudicated, the delivery encounters with a severe obstetric complication in the EHR data were shown to have a severe obstetric complication in the chart abstracted data, indicating strong measure validity. Although we do not always expect perfect agreement, as we expect some degree of human error in entering and matching values, we consider these PPV to show excellent measure score validity.
Sensitivity, specificity, and negative predictive value (NPV): Specificity and sensitivity results indicate high probability of the EHR data detecting a true severe obstetric complication during a delivery hospitalization based on the abstracted data ('gold standard'), and a high probability of the EHR data accurately identifying that no severe obstetric complication occurred during a delivery hospitalization. The strong NPV results indicate that when EHR data indicated a severe obstetric complication did not occur, and the chart abstraction confirmed a complication did not occur.
Overall, the study revealed ePC-07 to have an excellent measure outcome agreement rate of 91.2% with a kappa score of 0.881 indicating almost perfect agreement. An in-depth description of the findings is provided here:

- Pilot Site 1: 36 records across 10 individual hospitals exhibited a 97.2% measure outcome agreement rate with a kappa score of 0.963 indicating almost perfect agreement. Only one case resulted in a mismatched measure outcome. The ICD10 delivery code was missing from the procedure list and therefore the patient did not land in the initial population based on extracted data but in the denominator based on the adjudicated data.
- Pilot Site 2: 31 records for Pilot Site 2 exhibited a measure outcome agreement rate of 83.9% with kappa score of 0.786 indicating substantial agreement. In total, 5 cases mismatched. Two of the cases mismatched as the date assigned to the coded delivery procedure was incorrectly listed as the day before admission. Therefore, the patient was not in the initial population based on the raw data submitted by the hospital. The adjudicated data placed the patient in the denominator. The remaining 3 cases were in the denominator or stratum 1 based on

raw data. However, these 3 cases only met the initial population based on adjudicated data. This was due to the fact that the site submitted "null" for the Estimated Gestational Age (EGA) relevant date/time and EGA result. Estimated Date of Delivery (EDD) relevant date/time was reported as a date only. During reliability visit we found the EGA relevant date/time and EGA result; however, the EGA relevant date/time was after the date/time of delivery. We could not find the EDD relevant date/time; therefore, the adjudicated value was null. Neither the EDD nor EGA could be used to determine weeks gestation thereby rendering the patient ineligible for the denominator.

- Pilot Site 3: 35 records for Pilot Site 3 exhibited a measure outcome agreement rate of 94.3% with a kappa score of 0.922 indicating almost perfect agreement. Two of the cases mismatched based on a missing delivery time for one case and incorrect delivery procedure date. These errors resulted in patients qualifying for the initial population based on the original data and qualifying for the denominator based on the adjudicated data.
- Pilot Site 6: 36 records for Pilot Site 6 exhibited a measure outcome agreement rate of 97.2% with kappa score of 0.963 indicating almost perfect agreement. Only one case mismatched. The mismatch was due to an incorrect admission date/time which was after the delivery date/time. The correct admission time is prior to the delivery time and therefore the patient qualified for the denominator and not just the initial population as originally reported.
- Pilot Site 7: 30 records for Pilot Site 7 exhibited a measure outcome agreement rate of 96.7% with kappa score of 0.953 indicating near perfect agreement. The sole mismatch was due to a blood transfusion that was administered during a surgical procedure that was documented on paper and not available in a discrete field that could be reported from. The adjudicated data placed the patient in Stratum 1 instead of Stratum 2.
- Pilot Site 9: 36 records for Pilot Site 9 exhibited a measure outcome agreement rate of 77.8% with a kappa score of 0.703 indicating substantial agreement. Eight cases resulted in mismatched measure outcomes. Seven of the cases mismatched due to the fact the report writer extracted the baby's cord blood PaO2 level instead of the mother's PaO2 level. If not for this mapping error, the agreement rate would have been 97%. The site has identified how to rectify this error for future data pulls.

Data Element Agreement Rate Analysis:
Comment on feasibility scorecard in relationship to validity: As evidenced on the feasibility scorecards created for this measure, several data elements (PaO2/FiO2 result, POA indicator and procedure stop times) were scored as 0 for data accuracy by several hospitals. Validity testing proved that some of these data elements were problematic. Midway through validity testing, Joint Commission staff determined most pilot sites were unable to accurately capture the stop time for the procedure performed and the laboratory test result of the Pa02/Fi02 ratio. Joint Commission staff proposed to address these feasibility challenges by revising the draft specifications to better align with clinical intent and decrease burden for a lab result not commonly calculated in the EHR. As seen above in Table 2b.03.04, only 2 hospitals had less than perfect agreement rate on the Present On Admission indicator. See Pilot Site 1 and Pilot 6 analysis below for more details.
Overall, the data element agreement rate for all sites was excellent at a score of 90.4%. An in-depth description of the findings is provided here:

- Pilot Site 1 demonstrated a good agreement rate of 88%. CO2 results from chemistry panels were not reported for the bicarbonate data element. Present on admission codes of "U" (Unable to Determine) and "E" (Exempt) were mapped to "N" (No). Lastly, vital signs author date/times were used instead of assessment date/times. All these mapping errors are easily rectified.
- Pilot Site 2 demonstrated a fair data element agreement rate of 80.8%. This site had a substantial number of cases where massive blood protocol was initiated in the operating room. At that time, Anesthesia still documented on paper. Therefore, over 100 blood transfusions were not reported on the data file but were identified in the adjudicated data. As mentioned earlier, the EDD relevant date and time was only reported as a date. This could not be validated; therefore, the adjudicated data mismatched as null was reported upon reabstraction. EGA relevant date/time and EGA result were also problematic as previously described. While point of care testing glucose values were captured, glucose values resulted via chemistry panels were not. When perioperative vital signs were the first vital signs obtained, the report writer did not include them in the data extract. Workflows no longer segregate perioperative vital signs from all other vital signs.
- Pilot Site 3 demonstrated a near perfect data element agreement rate of 99.8%.
- Pilot Site 6 demonstrated a very good agreement rate of 91.3%. Very much like Pilot Site 1, Present on admission codes of "U" (Unable to Determine) and "E" (Exempt) were mapped to "N" (No) and CO2 results from chemistry panels were not reported for the bicarbonate data element.
- Pilot Site 7 demonstrated a good data element agreement rate of 86.3%. Dates and times of observation were not included in the data extract resulting in 87 mismatches. The admission times were mapped incorrectly by the

site. Had these dates and times been present and accurate, the overall agreement rate would have been 91.4%. Blood transfusion end times and laboratory results were problematic also with a 33% and 74% match rate respectively.

- Pilot Site 9 demonstrated an excellent agreement rate of 97.9%. Glucose results and resulted date/time were found to be mapped to an incorrect LOINC code not included in the Glucose value set. This resulted in some glucose values not appearing in the data submitted by the hospital. However, upon re-abstraction the glucose values were found in the EMR. The second area of concern was the inclusion of the baby's cord blood PaO2 level instead of the mother's PaO2 level as previously mentioned.

The face validity assessment demonstrated that the Technical Expert Panel members believe this eCQM is an important health outcome to measure because there is room for improvement, it will produce reliable and valid rates, and hospitals can use the results for performance improvement. While there are some concerns with the feasibility of implementation and whether this measure is a critical component of defining and comparing the quality of obstetric care between hospitals, the majority of the responses from the TEP either agreed or strongly agreed with the ability of this measure to improve patient outcomes.
The Patient Working Group members strongly believe this eCQM is an important health outcome to measure because there is room for improvement and strongly/moderately agree this measure is a critical component of defining and comparing the quality of obstetric care between hospitals.

[Response Ends]

2b.05. 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 in Importance to Measure and Report: Gap in Care/Disparities.

[Response Begins]
Variation in pilot site severe obstetric complication rates indicate a clinically meaningful quality gap in the delivery of maternal care to patients experiencing a delivery hospitalization, as some sites show results indicating higher rates of risk-standardized rates of severe obstetric complications while other sites show results indicating substantially lower risk-standardized rates of severe obstetric complications.
[Response Ends]

2b.06. Describe the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities.

Examples may include 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.

[Response Begins]
Table 2b.06.01 provides the observed and risk-standardized rate per 10,000 deliveries for severe obstetric complications and severe obstetric complications excluding blood transfusion only cases for each pilot site and across all sites. For the outcome of any severe obstetric complications, pilot site risk standardized results ranged from 158 delivery encounters with severe obstetric complications to 299 delivery encounters with severe obstetric complications. For the outcome of severe obstetric complications excluding blood transfusion-only cases, pilot site risk standardized results ranged from 48 to 55 delivery encounters with severe obstetric complications.
Table 2b.06.02 provides the observed and risk-standardized rate per 10,000 deliveries for severe obstetric complications and severe obstetric complications excluding blood transfusion only cases for each pilot hospital and across all hospitals. For the outcome of any severe obstetric complications, pilot hospital risk standardized results ranged from 157 delivery encounters with severe obstetric complications to 369 delivery encounters with severe obstetric complications. For the outcome of severe obstetric complications excluding blood transfusion-only cases, pilot hospital risk standardized results ranged from 49 to 55 delivery encounters with severe obstetric complications.
Please note there are minor discrepancies in the risk adjusted rates for the stand alone hospitals (2, 3, 6, 7, 9, 10) when comparing the 2 tables as the model was re-specified with a random effect component for 25 individual hospitals in Table 2b.06.02 instead of the original random effect component for the 8 sites in Table 2b.06.01. In adjusting for the 25

individual hospitals, it is expected that the predicted and expected SOC rates change a bit mathematically from the original model, and this leads to the slightly different measure scores.

Table 2b.06.01. Observed and Risk-Standardized Severe Obstetric Complication Rates per 10,000 Delivery Hospitalizations at the Site Level

Pilot Site	Delivery Encounters	Any Severe Obstetric Complication(s) Observed rate per 10,000 Delivery Hospitalizations	Any Severe Obstetric Complication(s) Risk-Standardized Rate per 10,000 Delivery Hospitalizations	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases Observed rate per 10,000 Delivery Hospitalizations	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases Risk-Standardized Rate per 10,000 Delivery Hospitalizations
Pilot Site 1	18,070	226	241	41	49
Pilot Site 2	7,196	235	248	72	55
Pilot Site 3	7,955	303	268	48	50
Pilot Site 5	6,139	209	223	44	50
Pilot Site 6	3,359	104	158	27	48
Pilot Site 7	4,369	213	255	41	50
Pilot Site 9	3,918	202	299	26	48
Pilot Site 10	9,178	341	285	81	51
Across Pilot Sites	60,184	244	252	50	50

Table 2b.06.01. displays the observed and risk-standardized severe obstetric complication (SOC) rates per 10,000 delivery hospitalizations at the Site Level. The number of delivery encounters and the observed rate of any SOC and SOC excluding blood transfusion-only cases are provided for each site. The risk-adjusted rates for any SOC and SOC excluding blood transfusion-only cases are also provided in the table. See additional details in prior paragraph.

Table 2b.06.02 Observed and Risk-Standardized Severe Obstetric Complication Rates per 10,000 Delivery Hospitalizations at the Hospital Level

Pilot Site	Delivery Encounters	Any Severe Obstetric Complication(s) Observed rate per 10,000 Delivery Hospitalizations	Any Severe Obstetric Complication(s) Risk-Standardized Rate per 10,000 Delivery Hospitalizations	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases Observed rate per 10,000 Delivery Hospitalizations	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases Risk-Standardized Rate per 10,000 Delivery Hospitalizations
1.1	**496**	**202**	**238**	**0**	**49**
1.2	**3875**	**248**	**284**	**52**	**51**
1.3	**1518**	**158**	**216**	**33**	**50**
1.4	**534**	**412**	**369**	**19**	**50**
1.5	**2383**	**105**	**163**	**29**	**50**
1.6	**5952**	**269**	**287**	**54**	**51**
1.7	**1678**	**244**	**315**	**36**	**50**
1.8	**733**	**164**	**209**	**14**	**50**
1.9	**608**	**214**	**223**	**16**	**49**
1.1	**293**	**171**	**233**	**34**	**50**
2	**7196**	**235**	**271**	**72**	**55**

Pilot Site	Delivery Encounters	Any Severe Obstetric Complication(s) Observed rate per 10,000 Delivery Hospitalizations	Any Severe Obstetric Complication(s) Risk-Standardized Rate per 10,000 Delivery Hospitalizations	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases Observed rate per 10,000 Delivery Hospitalizations	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases Risk-Standardized Rate per 10,000 Delivery Hospitalizations
3	7955	303	295	48	50
5.1	292	137	227	0	50
5.2	224	179	261	45	50
5.3	139	72	221	0	50
5.4	347	144	245	29	50
5.5	799	50	171	13	50
5.6	163	0	197	0	50
5.7	560	143	221	18	50
5.8	3316	305	295	66	51
5.9	299	33	187	33	50
6	3359	104	157	27	49
7	4369	213	281	41	50
9	3918	202	339	26	49
10	9178	341	314	81	51
Across Pilot Hospitals	60,184	244	249	50	50

Table 2b.06.02 displays the observed and risk-standardized severe obstetric complication rates per 10,000 delivery hospitalizations at the hospital level. Each pilot site is identified by the whole number and the individual hospitals within the pilot site are represented by the decimal. The number of delivery encounters for each hospital is provided in the table. See the prior paragraph for additional details.

[Response Ends]

2b.07. Provide 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.

In other words, what do the results mean in terms of statistical and meaningful differences?

[Response Begins]
The variation in severe obstetric complication rates suggests that there are meaningful differences in performance measure scores across pilot sites and hospitals. Variation in severe obstetric complication rates indicate a clinically meaningful quality gap in the delivery of maternal care to patients experiencing a delivery hospitalization, as some sites/hospitals show results indicating higher rates of risk-standardized rates of severe obstetric complications while other sites/hospitals show results indicating substantially lower risk-standardized rates of severe obstetric complications.
[Response Ends]

2b.08. Describe the method of testing conducted to identify the extent and distribution of missing data (or non-response) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders). Include how the specified handling of missing data minimizes bias.

Describe the steps—do not just name a method; what statistical analysis was used.

[Response Begins]
We developed this eCQM with the intent to, as much as possible, use variables that we expect to be consistently obtained in the target population, available in a structured field, and captured as part of standard clinical workflow. During feasibility testing, data elements were evaluated for feasibility and availability; two data elements were removed from measure specifications when several pilot sites were unable to accurately capture them (timestamp for procedure performed, and lab result for PaO2/FiO2 ratio). All other data elements were assessed to be feasible and available.
Many of the data elements used in the Severe Obstetric Complications eCQM are defined with ICD-10 diagnosis or procedure codes (for example, severe maternal mortality numerator events and risk adjustment variables). None of these data elements are considered to be missing when absent, since the absence of a given code implies absence of the corresponding condition.
For data elements representing vital signs and lab results, it is clinically acceptable that certain vital signs and labs were not performed for certain patients. That being said, vital sign and lab result fields with more than 20% missing were not considered as potential risk adjustment variables based on statistical considerations.

[Response Ends]

2b.09. Provide the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data.

For example, provide results of sensitivity analysis of the effect of various rules for missing data/non-response. If no empirical sensitivity analysis was conducted, identify the approaches for handling missing data that were considered and benefits and drawbacks of each).

[Response Begins]
Overall, the study revealed ePC-07 to have an excellent measure outcome agreement rate of 91.2% with a kappa score of 0.881 indicating almost perfect agreement. 2 pilot sites had mismatches due to missing data. Pilot Site 1 had only one case resulting in a mismatched measure outcome. The ICD10 delivery code was missing from the procedure list and therefore the patient did not land in the initial population based on extracted data but in the denominator based on the adjudicated data. Pilot Site 3 had one of the cases mismatched based on a missing delivery time. This error resulted in the patient qualifying for the initial population based on the original data and qualifying for the denominator based on the adjudicated data.
[Response Ends]

2b.10. Provide your interpretation of the results, in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and non-responders), and how the specified handling of missing data minimizes bias.

In other words, 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 was conducted, justify the selected approach for missing data.

[Response Begins]
As evidenced on the feasibility scorecards created for this measure, several data elements (PaO2/FiO2 result, POA indicator and procedure stop times) were scored as 0 for data accuracy by several hospitals. Validity testing proved that some of these data elements were problematic. Midway through validity testing, Joint Commission staff determined most pilot sites were unable to accurately capture the stop time for the procedure performed and the laboratory test result of the Pa02/Fi02 ratio. Joint Commission staff proposed to address these feasibility challenges by revising the draft specifications to better align with clinical intent and decrease burden for a lab result not commonly calculated in the EHR. As seen in Table 2b.03.04, only 2 hospitals had less than perfect agreement rate on the Present On Admission indicator. Please see 2b.04 for more details.

[Response Ends]

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 eCQMs). 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.

2b.11. Indicate whether there is more than one set of specifications for this measure.

[Response Begins]
No, there is only one set of specifications for this measure
[Response Ends]

2b.12. 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. Indicate what statistical analysis was used.

[Response Begins]
[Response Ends]

2b.13. Provide the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications.

Examples may include correlation, and/or rank order.

[Response Begins]
[Response Ends]

2b.14. Provide your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications.

In other words, what do the results mean and what are the norms for the test conducted.

[Response Begins]
[Response Ends]

2b.15. Indicate whether the measure uses exclusions.

[Response Begins]
Yes, the measure uses exclusions.
[Response Ends]

2b.16. Describe the method of testing exclusions and what was tested.

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?

[Response Begins]
We have compared the frequencies of the denominator and numerator by pilot site before and after the COVID exclusion. The performance scores were re-calculated and checked for any significant change after COVID exclusion. Since

the number of pilot sites is small, no formal statistical test has been performed for the effect of exclusion on the performance score.
[Response Ends]

2b.17. Provide 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.

[Response Begins]
Table 2b.17.01a Frequency Distribution of COVID Exclusions

Pilot Site Number	Denominator Delivery Encounters	Denominator Exclusions N	Denominator Exclusions %	Updated Denominator	Any Severe Obstetric Complication(s) Numerator	Any Severe Obstetric Complication(s) Exclusions N	Any Severe Obstetric Complication(s) Exclusions %	Any Severe Obstetric Complication(s) UPDATED	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases Numerator	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases Exclusions N	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases Exclusions %	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases UPDATED
1	18,070	20	0.11	18,050	408	6	1.47%	402	74	6	8.11%	68
2	7,196	3	0.04	7,193	169	1	0.59%	168	52	1	1.92%	51
3	7,955	6	0.08	7,949	241	4	1.66%	237	38	4	10.53%	34
5	6,139	6	0.10	6,133	128	1	0.78%	127	27	1	3.70%	26
6	3,359	0	0.00	3,359	35	0	0	35	9	0	0	9
7	4,369	1	0.02	4,368	93	0	0	93	18	0	0	18
9	3,918	1	0.03	3,917	79	1	1.27%	78	10	1	10%	9
10	9,178	0	0.00	9,178	313	0	0	313	74	0	0	74
Across	60,184	37	0.06	60,147	1,466	13	0.89%	1,453	302	13	4.30%	289

Table 2b.17.01a displays the frequency distribution of COVID exclusions across pilot sites. A total of 0.06% of denominator cases were excluded using the COVID exclusion criteria. Updated denominator and numerator rates are provided after exclusions as well as the updated Severe Obstetric Complication (SOC) numerators stratified by any SOC and SOC excluding blood transfusion-only numerator cases.
Table 2b.17.01b Frequency Distribution of COVID Exclusions by Hospital

Pilot Site Number	Denominator Delivery Encounters	Denominator Denominator Exclusions N	Denominator Denominator Exclusions %	Denominator Updated Denominator	Any Severe Obstetric Complication(s) Numerator	Any Severe Obstetric Complication(s) Exclusions N	Any Severe Obstetric Complication(s) Exclusions %	Any Severe Obstetric Complication(s) UPDATED	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases Numerator	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases Exclusions N	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases Exclusions %	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases UPDATED
1.1	496	0	0	496	10	0	0	10	0	0	*	0
1.2	3,875	5	0.13	3,870	96	3	3.13	93	20	3	15	17
1.3	1,518	1	0.07	1,517	24	0	0	24	5	0	0	5
1.4	534	0	0	534	22	0	0	22	1	0	0	1
1.5	2,383	4	0.17	2,379	25	1	4	24	7	1	14.29	6
1.6	5,952	6	0.1	5,946	160	2	1.25	158	32	2	6.25	30
1.7	1,678	4	0.24	1,674	41	0	0	41	6	0	0	6
1.8	733	0	0	733	12	0	0	12	1	0	0	1
1.9	608	0	0	608	13	0	0	13	1	0	0	1
1.1	293	0	0	293	5	0	0	5	1	0	0	1
2	7,196	3	0.04	7,193	169	1	0.59	168	52	1	1.92	51
3	7,955	6	0.08	7,949	241	4	1.66	237	38	4	10.53	34
5.1	292	1	0.34	291	4	0	0	4	0	0	*	0
5.2	224	0	0	224	4	0	0	4	1	0	0	1
5.3	139	1	0.72	138	1	0	0	1	0	0	*	0
5.4	347	0	0	347	5	0	0	5	1	0	0	1
5.5	799	0	0	799	4	0	0	4	1	0	0	1
5.6	163	0	0	163	0	0	*	0	0	0	*	0
5.7	560	0	0	560	8	0	0	8	1	0	0	1
5.8	3,316	4	0.12	3,312	101	1	0.99	100	22	1	4.55	21
5.9	299	0	0	299	1	0	0	1	1	0	0	1
6	3,359	0	0	3,359	35	0	0	35	9	0	0	9
7	4,369	1	0.02	4,368	93	0	0	93	18	0	0	18

Pilot Site Number	Denominator Delivery Encounters	Denominator Denominator Exclusions N	Denominator Denominator Exclusions %	Denominator Updated Denominator	Any Severe Obstetric Complication(s) Numerator	Any Severe Obstetric Complication(s) Exclusions N	Any Severe Obstetric Complication(s) Exclusions %	Any Severe Obstetric Complication(s) UPDATED	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases Numerator	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases Exclusions N	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases Exclusions %	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases UPDATED
9	3,918	1	0.03	3,917	79	1	1.27	78	10	1	10	9
10	9,178	0	0	9,178	313	0	0	313	74	0	0	74
Across	60,184	37	0.06	60,147	1,466	13	0.89	1,453	302	13	4.3	289

* This cell intentionally left empty.

Table 2b.17.01b displays the frequency distribution of COVID exclusions by hospitals. The range of denominator exclusions are from 0 to 6 cases per hospital. Updated denominator and numerator rates are provided after exclusions as well as the updated Severe Obstetric Complication (SOC) numerators stratified by any SOC and SOC excluding blood transfusion-only numerator cases.

Table 2b.17.02a COVID Exclusions Impact on performance measure scores

Pilot Site	Any Severe Obstetric Complication(s) Original Observed rate per 10,000 Delivery Hospitalizations	Any Severe Obstetric Complication(s) Original Risk-Standardized Rate per 10,000 Delivery Hospitalizations	Any Severe Obstetric Complication(s) With COVID Denominator Exclusion Observed rate per 10,000 Delivery Hospitalizations	Any Severe Obstetric Complication(s) With COVID Denominator Exclusion Risk-Standardized Rate per 10,000 Delivery Hospitalizations	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases Original Observed rate per 10,000 Delivery Hospitalizations	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases Risk-Standardized Rate per 10,000 Delivery Hospitalizations	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases With COVID Denominator Exclusion Observed rate per 10,000 Delivery Hospitalizations	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases With COVID Denominator Exclusion Risk-Standardized Rate per 10,000 Delivery Hospitalizations
1	226	241	223	236	41	49	38	45
2	235	248	234	246	72	55	71	58
3	303	268	298	262	48	50	43	45
5	209	223	207	223	44	50	42	48

Pilot Site	Any Severe Obstetric Complication(s) Original Observed rate per 10,000 Delivery Hospitalizations	Any Severe Obstetric Complication(s) Original Risk-Standardized Rate per 10,000 Delivery Hospitalizations Risk-Standardized Rate per 10,000 Delivery Hospitalizations	Any Severe Obstetric Complication(s) With COVID Denominator Exclusion Observed rate per 10,000 Delivery Hospitalizations	Any Severe Obstetric Complication(s) With COVID Denominator Exclusion Risk-Standardized Rate per 10,000 Delivery Hospitalizations	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases Original Observed rate per 10,000 Delivery Hospitalizations	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases Risk-Standardized Rate per 10,000 Delivery Hospitalizations	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases With COVID Denominator Exclusion Observed rate per 10,000 Delivery Hospitalizations	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases With COVID Denominator Exclusion Risk-Standardized Rate per 10,000 Delivery Hospitalizations
6	104	158	104	156	27	48	27	45
7	213	255	213	256	41	50	41	48
9	202	299	199	299	26	48	23	45
10	341	285	341	285	81	51	81	50
Across Sites	244	252	242	249	50	50	48	48

Table 2b.17.02a displays COVID exclusions' impact on performance measure scores at the pilot site level and across all sites.

Table 2b.17.02b COVID Exclusions Impact on performance measure scores, by individual hospital

Pilot Site	Any Severe Obstetric Complication(s) Original Observed rate per 10,000 Delivery Hospitalizations	Any Severe Obstetric Complication(s) Original Risk-Standardized Rate per 10,000 Delivery Hospitalizations Risk-Standardized Rate per 10,000 Delivery Hospitalizations	Any Severe Obstetric Complication(s) With COVID Denominator Exclusion Observed rate per 10,000 Delivery Hospitalizations	Any Severe Obstetric Complication(s) With COVID Denominator Exclusion Risk-Standardized Rate per 10,000 Delivery Hospitalizations	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases Original Observed rate per 10,000 Delivery Hospitalizations	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases Risk-Standardized Rate per 10,000 Delivery Hospitalizations	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases With COVID Denominator Exclusion Observed rate per 10,000 Delivery Hospitalizations	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases With COVID Denominator Exclusion Risk-Standardized Rate per 10,000 Delivery Hospitalizations
1.1	202	238	202	236	0	49	0	47
1.2	248	284	240	274	52	51	44	49

Pilot Site	Any Severe Obstetric Complication(s) Original Observed rate per 10,000 Delivery Hospitalizations	Any Severe Obstetric Complication(s) Original Risk-Standardized Rate per 10,000 Delivery Hospitalizations	Any Severe Obstetric Complication(s) With COVID Denominator Exclusion Observed rate per 10,000 Delivery Hospitalizations	Any Severe Obstetric Complication(s) With COVID Denominator Exclusion Risk-Standardized Rate per 10,000 Delivery Hospitalizations	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases Original Observed rate per 10,000 Delivery Hospitalizations	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases Risk-Standardized Rate per 10,000 Delivery Hospitalizations	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases With COVID Denominator Exclusion Observed rate per 10,000 Delivery Hospitalizations	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases With COVID Denominator Exclusion Risk-Standardized Rate per 10,000 Delivery Hospitalizations
1.3	158	216	158	214	33	50	33	48
1.4	412	369	412	369	19	50	19	47
1.5	105	163	101	158	29	50	25	47
1.6	269	287	266	283	54	51	50	49
1.7	244	315	245	314	36	50	36	48
1.8	164	209	164	207	14	50	14	47
1.9	214	223	214	221	16	49	16	47
1.1	171	233	171	230	34	50	34	48
2	235	271	234	269	72	55	71	56
3	303	295	298	289	48	50	43	46
5.1	137	227	137	226	0	50	0	47
5.2	179	261	179	260	45	50	45	48
5.3	72	221	72	219	0	50	0	48
5.4	144	245	144	244	29	50	29	48
5.5	50	171	50	170	13	50	13	48
5.6	0	197	0	194	0	50	0	48
5.7	143	221	143	220	18	50	18	48
5.8	305	295	302	296	66	51	63	50
5.9	33	187	33	185	33	50	33	48
6	104	157	104	156	27	49	27	46
7	213	281	213	283	41	50	41	48
9	202	339	199	339	26	49	23	46
10	341	314	341	314	81	51	81	50
Across Sites	244	249	242	275	50	50	48	48

Table 2b.17.02b displays the COVID exclusions' impact on performance measure scores for individual hospitals and across all sites.

[Response Ends]

2b.18. Provide your interpretation of the results, in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results.

In other words, 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.

[Response Begins]
The evidence base for COVID-19 and related variants is rapidly growing and changing. Available studies suggest that symptomatic pregnant women with COVID-19 are at increased risk of more severe illness compared with nonpregnant peers (Kahn, 2021). The COVID-19 exclusion was added to ensure patients with this condition who were symptomatic with respiratory conditions would not be counted as a numerator case for hospitals. Although rare, cases with this exclusion were found during analysis. Treatment protocols are being developed and tested and the measure should not include these patients while preventability of these complications is unknown.
Khan, D., Pirzada, A. N., Ali, A., Salam, R. A., Das, J. K., & Lassi, Z. S. (2021). The Differences in Clinical Presentation, Management, and Prognosis of Laboratory-Confirmed COVID-19 between Pregnant and Non-Pregnant Women: A Systematic Review and Meta-Analysis. *International journal of environmental research and public health*, 18(11), 5613. https://doi.org/10.3390/ijerph18115613

[Response Ends]

2b.19. Check all methods used to address risk factors.

[Response Begins]
Statistical risk model with risk factors (specify number of risk factors)
[Statistical risk model with risk factors (specify number of risk factors) Please Explain]
34- See 2b.20 for additional details.

[Response Ends]

2b.20. If using statistical risk models, provide detailed risk model specifications, including the risk model method, risk factors, risk factor data sources, coefficients, equations, codes with descriptors, and definitions.

[Response Begins]
Risk model development performed by Yale New Haven Health Services Corporation/Center for Outcomes Research and Evaluation (CORE).
Following the identification of risk-adjustment variables (described in 2b.23 below), a risk model was developed for both outcomes: severe obstetric complications and severe obstetric complications excluding blood transfusion-only encounters. The risk model was developed and tested with data from eight pilot sites; 60,184 delivery hospitalizations were randomly divided in a 70/30 split for a development dataset (N=42,129) and a validation dataset (N=18,055).
The following variables were included in the final risk model:

- Demographics and patient characteristics: maternal age
- Preexisting conditions and pregnancy characteristics defined by ICD-10 codes
- Anemia
- Asthma
- Autoimmune disease
- Bariatric surgery
- Bleeding disorder
- Body Mass Index (BMI)
- Cardiac disease
- Gastrointestinal disease

- Gestational diabetes
- Human Immunodeficiency Virus (HIV)
- Hypertension
- Mental health disorder
- Multiple pregnancy
- Neuromuscular disease
- Obstetric venous thromboembolism (VTE)
- Other pre-eclampsia
- Placental accreta spectrum
- Placental abruption
- Placenta previa
- Preexisting diabetes
- Preterm birth
- Previous cesarean
- Pulmonary hypertension
- Renal disease
- Severe pre-eclampsia
- Substance abuse
- Thyrotoxicosis
- Laboratory tests and vital signs upon hospital arrival (Hematocrit, White blood cell [WBC] count, Heart rate, Systolic blood pressure)
- Long-term anticoagulant medication use
- Social Risk Factors: economic/housing instability

With the list of risk variables identified for the risk model, we estimated the hospital-specific risk standardized obstetric complications rate (RSOCR) using a hierarchical logistic regression model (hierarchical model). This strategy accounts for within-hospital correlation of the observed outcome among patients and accommodates the assumption that underlying differences in the quality of care across hospitals lead to systematic differences in patient outcomes. This approach models the log odds of a severe obstetric complication as a function of patient demographics and clinically relevant comorbidities with a random intercept for the hospital-specific effect.

The hospital-specific RSOCRs were calculated as the ratio of a hospital's "predicted" number of delivery hospitalizations with a severe obstetric complication to "expected" number of delivery hospitalizations with a severe obstetric complication multiplied by the overall observed rate of delivery hospitalizations with a severe obstetric complication. The expected number of delivery hospitalizations with a complication for each hospital (denominator) was estimated using its patient mix and the average hospital-specific intercept (i.e., the average intercept among all hospitals in the sample). The predicted number of delivery hospitalizations with a complication for each hospital (numerator) was estimated given the same patient mix but an estimated hospital-specific intercept. Operationally, the expected number of delivery hospitalizations with a complication for each hospital was obtained by summing the expected complications for all delivering patients in the hospital. The expected complications outcome for each delivering patient was calculated via the hierarchical model, which applies the estimated regression coefficients to the observed patient characteristics and adds the average of the hospital-specific intercept. The predicted number of delivery hospitalizations with a complication for each hospital was calculated by summing the predicted complications for all delivering patients in the hospital. The predicted complications outcome for each delivering patient was calculated through the hierarchical model, which applies the estimated regression coefficients to the patient characteristics observed and adds the hospital-specific intercept.

More specifically, we used a hierarchical model to account for the natural clustering of observations within hospitals. The model employs a logit link function to link the risk factors to the outcome with a hospital-specific random effect:

- Let Y_{ij} denote the outcome (equal to one if the delivery encounter has a severe obstetric complication, zero otherwise) for patient *i* at hospital *j*; Z_{ij} denotes a set of risk factors for patient *i* at hospital *j*; and n_j is the number of delivery admissions to hospital *j*. We assume the outcome is related linearly to the covariates via a logit function:

- Logistic Regression Model
  - $$\boldsymbol{logit(Prob(Y_{ij} = 1)) = \alpha + \beta Z_{ij}} \quad (1)$$
- and $\boldsymbol{Z_{ij} = (Z_{1ij}, Z_{2ij}, \dots, Z_{pij})}$ is a set of p patient-specific covariates.

- To account for the natural clustering of observations within hospitals, we estimate a hierarchical logistic regression model that links the risk factors to the same outcomes and a hospital-specific random effect.

- Hierarchical Logistic Regression Model
  - $\boldsymbol{logit(Prob(Y_{ij} = 1)) = a_j + \beta Z_{ij}}$ (2)
  - where $\boldsymbol{\alpha_j = \mu + \omega_j; \omega_j\ N(0, \tau^2)}$ (3)

- where α_j represents the hospital-specific intercept, Z_{ij} is defined as above, μ is the adjusted average intercept over all hospitals in the sample, ω_j is the hospital-specific intercept deviation from μ, and $2\tau^2$ is the between-hospital variance component. This model separates within-hospital variation from between-hospital variation. Both the hierarchical logistic regression model and the logistic regression model are estimated using the SAS software system (GLIMMIX and LOGISTIC procedures, respectively).

- The risk model detail, including coefficients, is given in Table 2b.20.01. This table provides frequencies and adjusted odds ratios (ORs) and 95% confidence intervals (CIs) from the hierarchical model for the final set of demographic and clinical variables used for risk adjustment.

Table 2b.20.01 Risk Variables w/Adjusted Odds Ratio (OR) for Risk Model for Delivery Hospitalizations with Any Severe Obstetric Complication(s) and Risk Model of Delivery Hospitalizations with Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases

Variable	Full Sample = 60,184 n(%)	Adjusted OR (95% CI) Any SMM	Adjusted OR (95% CI) Any SMM Excluding Blood Transfusion-Only Cases	Beta Coefficients (Standard Error) Any SMM	Beta Coefficients (Standard Error) Any SMM Excluding Blood Transfusion-Only Cases
Average Age in Years [Mean (STD)]	30 (6)	N/A	N/A	N/A	N/A
<20	1,574 (2.6%)	REF	REF	REF	REF
20-<25	8,558 (14.2%)	1.05 (0.76, 1.45)	1.01 (0.42, 2.44)	0.04 (0.16)	0.01 (0.45)
25-<30	15,512 (25.8%)	0.85 (0.62, 1.16)	1.24 (0.53, 2.90)	-0.17 (0.16)	0.21 (0.43)
30-<35	20,237 (33.6%)	0.83 (0.60, 1.13)	1.26 (0.54, 2.93)	-0.20 (0.16)	0.22 (0.43)
35-<40	11,499 (19.1%)	0.86 (0.62, 1.19)	1.07 (0.45, 2.54)	-0.16 (0.17)	0.06 (0.44)
>=40	2,804 (4.7%)	1.41 (0.97, 2.03)	1.92 (0.76, 4.87)	0.33 (0.19)	0.65 (0.47)
Anemia	11,466 (19.1%)	1.76 (1.56, 1.98)	1.45 (1.10, 1.92)	0.56 (0.06)	0.37 (0.14)
Asthma	5,099 (8.5%)	1.21 (1.02, 1.43)	2.00 (1.46, 2.73)	0.19 (0.09)	0.69 (0.16)
Autoimmune Disease	157 (0.3%)	2.21 (1.16, 4.23)	NA*	0.79 (0.33)	NA*
BMI	12,047 (20.0%)	1.04 (0.91, 1.20)	1.21 (0.90, 1.61)	0.04 (0.07)	0.19 (0.15)
Bariatric Surgery	445 (0.7%)	0.93 (0.54, 1.60)	0.80 (0.24, 2.68)	-0.07 (0.27)	-0.22 (0.62)

Variable	Full Sample = 60,184 n(%)	Adjusted OR (95% CI) Any SMM	Adjusted OR (95% CI) Any SMM Excluding Blood Transfusion-Only Cases	Beta Coefficients (Standard Error) Any SMM	Beta Coefficients (Standard Error) Any SMM Excluding Blood Transfusion-Only Cases
Bleeding Disorder	1,768 (2.9%)	2.09 (1.66, 2.62)	2.50 (1.62, 3.87)	0.74 (0.12)	0.92 (0.22)
Cardiac Disease	939 (1.6%)	1.61 (1.18, 2.18)	2.86 (1.74, 4.70)	0.47 (0.16)	1.05 (0.25)
Economic_Housing Instability	62 (0.1%)	1.79 (0.66, 4.85)	5.10 (1.44, 18.10)	0.58 (0.51)	1.63 (0.65)
Gastrointestinal Disease	967 (1.6%)	1.28 (0.90, 1.81)	1.01 (0.47, 2.19)	0.24 (0.18)	0.01 (0.39)
Gestational Diabetes	5,793 (9.6%)	1.04 (0.87, 1.24)	1.43 (1.02, 2.02)	0.04 (0.09)	0.36 (0.18)
HIV	71 (0.1%)	1.75 (0.69, 4.49)	NA*	0.56 (0.48)	NA*
Hypertension	2,613 (4.3%)	0.99 (0.79, 1.24)	0.77 (0.48, 1.23)	-0.01 (0.11)	-0.26 (0.24)
Long Term Anticoagulant Use	181 (0.3%)	1.26 (0.66, 2.42)	NA*	0.23 (0.33)	NA*
Mental Health Disorder	8,753 (14.5%)	1.23 (1.07, 1.41)	1.27 (0.95, 1.71)	0.21 (0.07)	0.24 (0.15)
Multiple Pregnancy	1,178 (2.0%)	2.11 (1.64, 2.70)	1.48 (0.84, 2.60)	0.75 (0.13)	0.39 (0.29)
Neuromuscular	303 (0.5%)	0.94 (0.47, 1.87)	0.98 (0.23, 4.13)	-0.06 (0.35)	-0.02 (0.73)
Obstetrical VTE	52 (0.1%)	0.58 (0.11, 2.94)	NA*	-0.54 (0.83)	NA*
Other Preeclampsia	6,025 (10.0%)	1.32 (1.11, 1.56)	1.44 (0.99, 2.11)	0.28 (0.09)	0.37 (0.19)
Placenta Previa	271 (0.5%)	3.94 (2.60, 5.95)	1.36 (0.58, 3.18)	1.37 (0.21)	0.31 (0.43)
Placental Abruption	548 (0.9%)	3.69 (2.76, 4.93)	2.52 (1.32, 4.79)	1.31 (0.15)	0.92 (0.33)
Placental Accreta Spectrum	66 (0.1%)	50.11 (27.20, 92.32)	174.25 (91.18, 333.00)	3.91 (0.31)	5.16 (0.33)
Preexisting Diabetes	903 (1.5%)	1.61 (1.19, 2.19)	1.91 (1.11, 3.28)	0.48 (0.16)	0.64 (0.28)
Preterm Birth	4,097 (6.8%)	1.37 (1.15, 1.63)	2.22 (1.59, 3.09)	0.31 (0.09)	0.80 (0.17)
Previous Cesarean	10,256 (17.0%)	1.29 (1.13, 1.48)	1.15 (0.85, 1.55)	0.26 (0.07)	0.14 (0.15)
Pulmonary Hypertension	23 (0.0%)	0.99 (0.23, 4.24)	3.23 (0.76, 13.65)	-0.01 (0.74)	1.17 (0.74)

Variable	Full Sample = 60,184 n(%)	Adjusted OR (95% CI) Any SMM	Adjusted OR (95% CI) Any SMM Excluding Blood Transfusion-Only Cases	Beta Coefficients (Standard Error) Any SMM	Beta Coefficients (Standard Error) Any SMM Excluding Blood Transfusion-Only Cases
Renal Disease	146 (0.2%)	2.80 (1.68, 4.69)	3.13 (1.41, 6.94)	1.03 (0.26)	1.14 (0.41)
Severe Preeclampsia	2,337 (3.9%)	2.56 (2.07, 3.16)	3.92 (2.62, 5.87)	0.94 (0.11)	1.37 (0.21)
Substance Abuse	4,048 (6.7%)	1.06 (0.88, 1.27)	1.21 (0.81, 1.79)	0.05 (0.10)	0.19 (0.20)
Thyrotoxicosis	212 (0.4%)	0.41 (0.13, 1.31)	0.67 (0.09, 4.91)	-0.90 (0.60)	-0.40 (1.01)
Low prevalence factors 1: Autoimmune disease OR HIV	227 (0.4%)	NA	1.67 (0.51, 5.54)	NA	0.52 (0.61)
Low prevalence factors 2: Long Term Anticoagulant Use OR Obstetrical VTE	224 (0.4%)	NA	0.95 (0.30, 2.99)	NA	-0.05 (0.59)
Vitals - Heart Rate	*	*	*	*	*
Result <110	50,945 (84.6%)	REF	REF	REF	REF
Result >=110	5,607 (9.3%)	1.25 (1.06, 1.48)	1.41 (0.99, 2.00)	0.23 (0.09)	0.34 (0.18)
Missing	3,632 (6.0%)	2.32 (1.23, 4.40)	1.77 (0.37, 8.58)	0.84 (0.33)	0.57 (0.80)
Vitals - Systolic BP	*	*	*	*	*
Result <140	47,677 (79.2%)	REF	REF	REF	REF
Result >=140 & <160	7,275 (12.1%)	1.11 (0.94, 1.31)	0.95 (0.67, 1.36)	0.11 (0.08)	-0.05 (0.18)
Result >=160	1,664 (2.8%)	1.13 (0.87, 1.48)	0.61 (0.34, 1.09)	0.13 (0.14)	-0.49 (0.29)
Missing	3,568 (5.9%)	0.65 (0.33, 1.28)	0.91 (0.18, 4.64)	-0.43 (0.34)	-0.09 (0.83)
Labs - Hematocrit	*	*	*	*	*
Result <33	11,344 (18.8%)	2.66 (2.36, 3.01)	1.13 (0.84, 1.53)	0.98 (0.06)	0.12 (0.15)
Result >=33	41,293 (68.6%)	REF	REF	REF	REF
Missing	7,547 (12.5%)	1.25 (0.95, 1.66)	0.82 (0.45, 1.49)	0.23 (0.14)	-0.19 (0.30)
Labs - WBC	*	*	*	*	*
Result <14	42,099 (70.0%)	REF	REF	REF	REF

Variable	Full Sample = 60,184 n(%)	Adjusted OR (95% CI) Any SMM	Adjusted OR (95% CI) Any SMM Excluding Blood Transfusion-Only Cases	Beta Coefficients (Standard Error) Any SMM	Beta Coefficients (Standard Error) Any SMM Excluding Blood Transfusion-Only Cases
Result >=14	7,010 (11.6%)	1.19 (1.01, 1.40)	1.46 (1.05, 2.04)	0.18 (0.08)	0.38 (0.17)
Missing	11,075 (18.4%)	0.60 (0.47, 0.76)	0.68 (0.41, 1.13)	-0.52 (0.12)	-0.39 (0.26)

* Due to low prevalence of select risk variables, for the risk model of severe obstetric complication excluding transfusion-only cases, Human Immunodeficiency Virus (HIV) was combined with autoimmune disease, and obstetric venous thromboembolism (VTE) was combined with long-term anticoagulant medication use

Table 2b.20.01 displays risk variables w/adjusted Odds Ratio (OR) and Beta Coefficients for risk model for delivery hospitalizations with any Severe Obstetric Complication(s) (SOC) and risk model of delivery hospitalizations with Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases. The full sample n and the n for each individual variable are provided.

[Response Ends]

2b.21. If an outcome or resource use measure is not risk-adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (i.e., case mix) is not needed to achieve fair comparisons across measured entities.

[Response Begins]
[Response Ends]

2b.22. Select all applicable resources and methods used to develop the conceptual model of how social risk impacts this outcome.

[Response Begins]
Published literature
Internal data analysis
[Response Ends]

2b.23. Describe the conceptual and statistical methods and criteria used to test and select patient-level risk factors (e.g., clinical factors, social risk factors) used in the statistical risk model or for stratification by risk.

Please be sure to address the following: potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10 or other statistical tests; correlation of x or higher. Patient factors should be present at the start of care, if applicable. Also discuss any "ordering" of risk factor inclusion; note whether social risk factors are added after all clinical factors. Discuss any considerations regarding data sources (e.g., availability, specificity).

[Response Begins]
We identified candidate risk variables for SMM for consideration in the measure risk adjustment model by utilizing literature and research findings, including An Expanded Obstetric Comorbidity Scoring System for Predicting Severe Maternal Morbidity by Dr. Stephanie Leonard, the NQF Maternal Morbidity and Mortality Environmental Scan, and our initial ES/LR findings on specific drivers of severe obstetric complications and maternal mortality (Leonard et al., 2020; National Quality Forum, 2020). We also solicited input from clinicians, patients, and other experts in the TEP who identified for consideration numerous risk-adjustment variables at the patient and hospital levels. These included, but were not limited to, prior pregnancy history, housing instability, and availability of specialists and trauma care in hospitals. The teams acknowledged and carefully considered recommendations from the TEP and Patient Working Group for selection of candidate risk-adjustment variables.

Risk variables were removed from inclusion in the model if there were greater than 20% missing values (relevant for vital signs and laboratory results). In addition, due to a lack of variation across encounters, temperature and respiratory rate were not included in the final model. The same risk variables were included in the risk models for severe obstetric complications and severe obstetric complications excluding blood transfusion-only encounters; however, due to very low prevalence of a few risk variables in the risk model of severe obstetric complication excluding transfusion-only cases, Human Immunodeficiency Virus (HIV) was combined with autoimmune disease, and obstetric venous thromboembolism (VTE) was combined with long-term anticoagulant medication use.

Leonard, S.A., Kennedy, C.J., Carmichael, S.L., Lyell, D.J., & Main, E.K. (2020). An expanded obstetric comorbidity scoring system for predicting severe maternal morbidity. *Obstetrics & Gynecology*, 136(3), 440-449.
National Quality Forum (2020). Maternal Morbidity and Mortality Environmental Scan.

[Response Ends]

2b.24. Detail the statistical results of the analyses used to test and select risk factors for inclusion in or exclusion from the risk model/stratification.

[Response Begins]
Please refer to section 2b.20. Table 2b20.01 provides frequencies and adjusted odds ratios (ORs) and 95% confidence intervals (CIs) from the hierarchical model for the final set of demographic and clinical variables used for risk adjustment.
[Response Ends]

2b.25. Describe the analyses and interpretation resulting in the decision to select or not select social risk factors.

Examples may include prevalence of the factor across measured entities, availability of the data source, empirical association with the outcome, contribution of unique variation in the outcome, or assessment of between-unit effects and within-unit effects. Also describe the impact of adjusting for risk (or making no adjustment) on providers at high or low extremes of risk.

[Response Begins]
The decisions to include housing/economic instability as a risk factor and race/ethnicity as a stratification factor were made a priori and were not tested or influenced by analytic results.
Our goal in selecting risk factors for adjustment was to develop parsimonious models that included clinically relevant variables strongly associated with a severe obstetric complication outcome. We used a two-stage approach, first identifying the comorbidity or clinical status risk factors that were most important in predicting the outcome, then considering the potential addition of social risk factors. Social risk factors considered were also dependent on the availability of information in the EHR. As noted above, economic/housing instability was included in the model, and was chosen due to support in research literature for its inclusion and availability in the EHR.
Because of the stark differences in maternal outcomes by race/ethnicity as demonstrated in the literature, these social risk factors were examined as stratification variables rather than risk variables, as discussed below. It was determined that illumination of outcome disparities by race/ethnicity, rather than adjustment of outcomes by race/ethnicity, would best inform stakeholders and patients and be most impactful in incentivizing improvements in quality of maternal care.

[Response Ends]

2b.26. 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 control for differences in patient characteristics (i.e., case mix) below. If stratified ONLY, enter "N/A" for questions about the statistical risk model discrimination and calibration statistics.

Validation testing should be conducted in a data set that is separate from the one used to develop the model.

[Response Begins]

To assess model performance, we computed discrimination and calibration statistics for assessing model performance (Harrell & Shih, 2001) for the clinically derived models, including:
(1) Area under the receiver operating characteristic (ROC) curve (the c-statistic [also called ROC] is the probability that predicting the outcome is better than chance, which is a measure of how accurately a statistical model can distinguish between a patient with and without an outcome)
(2) Predictive ability (discrimination in predictive ability measures the ability to distinguish high-risk subjects from low-risk subjects; good discrimination indicated by a wide range between the lowest decile and highest decile)
(3) Over-fitting indices (over-fitting refers to the phenomenon in which a model accurately describes the relationship between predictive variables and outcome in the development dataset but fails to provide valid predictions in new patients). A value of close to zero for the intercept and close to 1 for coefficient of risk score indicates good calibration of the model.
Results of model performance analyses are provided in 2b.27 and 2b.28, below.
Table 2b.26.01 and Table 2b.26.02 provides the observed and the risk-standardized rate per 10,000 deliveries rates for severe obstetric complications and severe obstetric complications excluding blood transfusion-only cases for each pilot site and at the hospital level, respectively. Please note there are minor discrepancies in the risk adjusted rates for the stand alone hospitals (2, 3, 6, 7, 9, 10) when comparing the 2 tables as the model was re-specified with a random effect component for 25 individual hospitals in Table 2b.26.02 instead of the original random effect component for the 8 sites in Table 2b.26.01. In adjusting for the 25 individual hospitals, it is expected that the predicted and expected SOC rates change a bit mathematically from the original model, and this leads to the slightly different measure scores.
Harrell, F.E., Shih, Y-C.T. (2001). Using full probability models to compute probabilities of actual interest to decision makers. *International journal of technology assessment in health care*, 17(1),17-26 .

Table 2b.26.01 Observed and Risk-Standardized Severe Obstetric Complication Rates per 10,000 Delivery Hospitalizations at the Site Level

Pilot Site	Delivery Encounters	Any Severe Obstetric Complication(s) Observed rate per 10,000 Delivery Hospitalizations	Any Severe Obstetric Complication(s) Risk-Standardized Rate per 10,000 Delivery Hospitalizations	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases Observed rate per 10,000 Delivery Hospitalizations	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases Risk-Standardized Rate per 10,000 Delivery Hospitalizations
Pilot Site 1	18070	226	241	41	49
Pilot Site 2	7196	235	248	72	55
Pilot Site 3	7955	303	268	48	50
Pilot Site 5	6139	209	223	44	50
Pilot Site 6	3359	104	158	27	48
Pilot Site 7	4369	213	255	41	50
Pilot Site 9	3918	202	299	26	48
Pilot Site 10	9178	341	285	81	51
Across Pilot Sites	60184	244	252	50	50

Table 2b.26.01 displays the observed and risk-standardized Severe Obstetric Complication (SOC) rates per 10,000 delivery hospitalizations at the pilot site level. Delivery encounters are provided for each site and the total number of delivery encounters is 60184. The SOC rates are provided for any SOC and SOC excluding blood transfusion-only cases.

Table 2b.26.02 Observed and Risk-Standardized Severe Obstetric Complication Rates per 10,000 Delivery Hospitalizations at the Hospital Level

Pilot Site	Delivery Encounters	Any Severe Obstetric Complication(s) Observed rate per 10,000 Delivery Hospitalizations	Any Severe Obstetric Complication(s) Risk-Standardized Rate per 10,000 Delivery Hospitalizations	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases Observed rate per 10,000 Delivery Hospitalizations	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases Risk-Standardized Rate per 10,000 Delivery Hospitalizations
1.1	496	202	238	0	49
1.2	3875	248	284	52	51
1.3	1518	158	216	33	50
1.4	534	412	369	19	50
1.5	2383	105	163	29	50
1.6	5952	269	287	54	51
1.7	1678	244	315	36	50
1.8	733	164	209	14	50
1.9	608	214	223	16	49
1.1	293	171	233	34	50
2	7196	235	271	72	55
3	7955	303	295	48	50
5.1	292	137	227	0	50
5.2	224	179	261	45	50
5.3	139	72	221	0	50
5.4	347	144	245	29	50
5.5	799	50	171	13	50
5.6	163	0	197	0	50
5.7	560	143	221	18	50
5.8	3316	305	295	66	51
5.9	299	33	187	33	50
6	3359	104	157	27	49
7	4369	213	281	41	50
9	3918	202	339	26	49
10	9178	341	314	81	51
Across Pilot Hospitals	60,184	244	249	50	50

Table 2b.26.02 displays the observed and risk-standardized Severe Obstetric Complication (SOC) rates per 10,000 delivery hospitalizations at the hospital level. Each individual hospitals number of delivery encounters is provided. The SOC rates are provided for any SOC and SOC excluding blood transfusion-only cases.

[Response Ends]

2b.27. Provide risk model discrimination statistics.

For example, provide c-statistics or R-squared values.

[Response Begins]

Table 2b.27.01 provides C-statistic and predictability results.
The calculated C-statistic for the risk model for any severe obstetric complications was 0.74 using the development dataset and 0.75 using the validation dataset; the calculated C-statistic for the severe obstetric complications excluding blood transfusion-only cases measure was 0.77 using the development dataset and 0.73 using the validation dataset. With both the development and validation datasets, both models show a reasonable range between the lowest decile and highest decile of predicted ability, given the low prevalence of the outcome. Overall, these diagnostic results demonstrate the risk-adjustment model adequately controls for differences in patient characteristics.

Table 2b.27.01 Model Performance Statistics for Risk Model for Delivery Hospitalizations with Any Severe Obstetric Complication(s) and Risk Model of Delivery Hospitalizations with Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases

Model Performance Statistic	Any Severe Obstetric Complication(s) Development Dataset	Any Severe Obstetric Complication(s) Validation Dataset	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases Development Dataset	Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases Validation Dataset
C-statistic	0.74 (0.72,0.76)	0.75 (0.72,0.77)	0.77 (0.73,0.81)	0.73 (0.67,0.80)
Predictability	(0.72,9.63)	(0.45,10.07)	(0.17,2.59)	(0.12,2.49)

Table 2b.27.01 displays the model performance statistics for the risk model for delivery hospitalizations with Any Severe Obstetric Complication(s) (SOC) and risk model of delivery hospitalizations with Severe Obstetric Complication(s) Excluding Blood Transfusion-Only Cases. The C-statistic and Predictability are provided for both the development and validation dataset.

[Response Ends]

2b.28. Provide the statistical risk model calibration statistics (e.g., Hosmer-Lemeshow statistic).

[Response Begins]
The calibration indices (γ0, γ1) used to assess the risk model for the any severe obstetric complications in the validation dataset are (0.15, 1.05) and for the severe obstetric complications excluding blood transfusion-only cases in the validation dataset are (0.22, 1.04).
[Response Ends]

2b.29. Provide the risk decile plots or calibration curves used in calibrating the statistical risk model.

The preferred file format is .png, but most image formats are acceptable.

[Response Begins]
Figure 2b.29.01 Internal Calibration for Rates of Any Severe Obstetric Complication

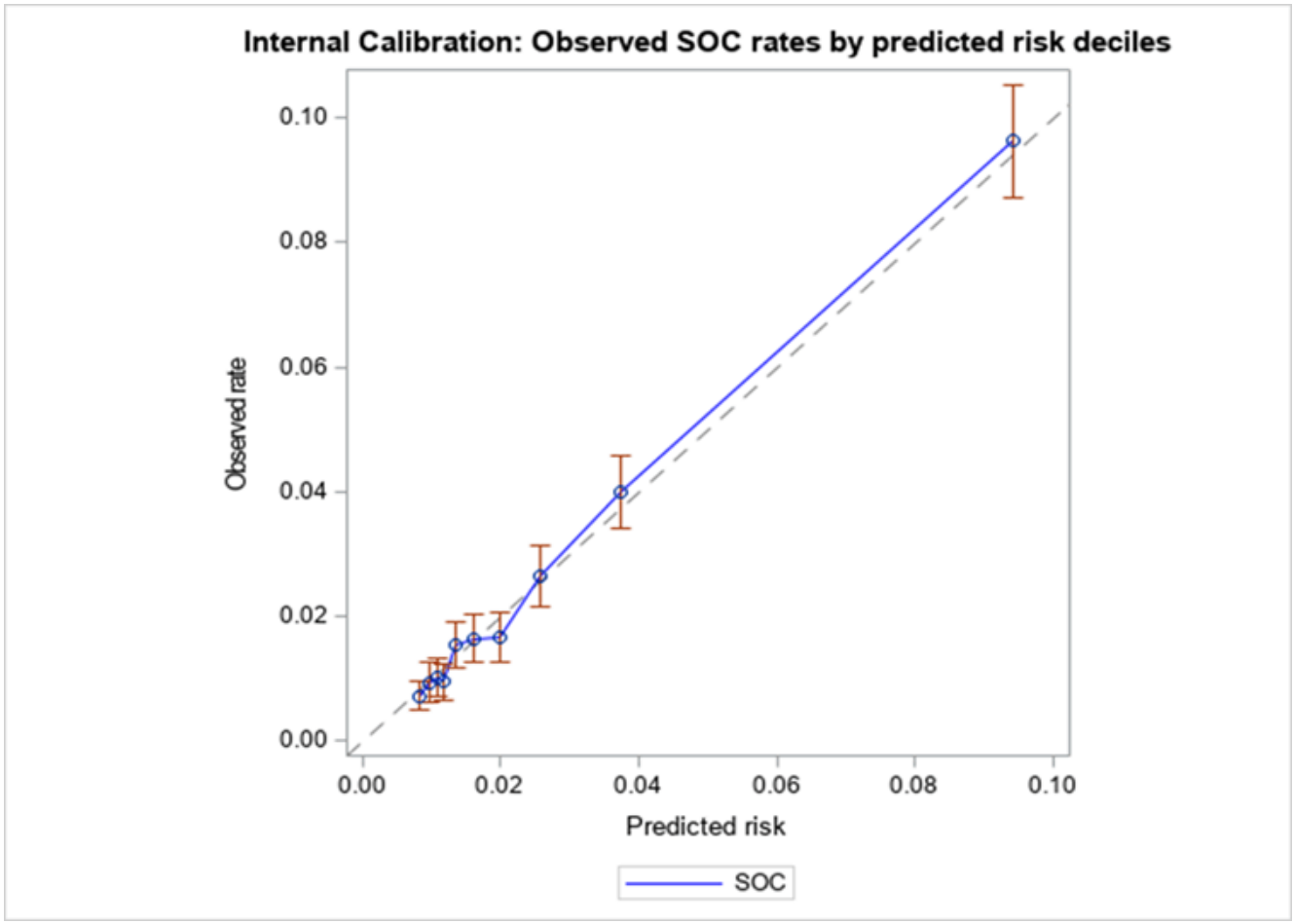


Figure 2b.29.02 External Calibration for Rates of Any Severe Obstetric Complication

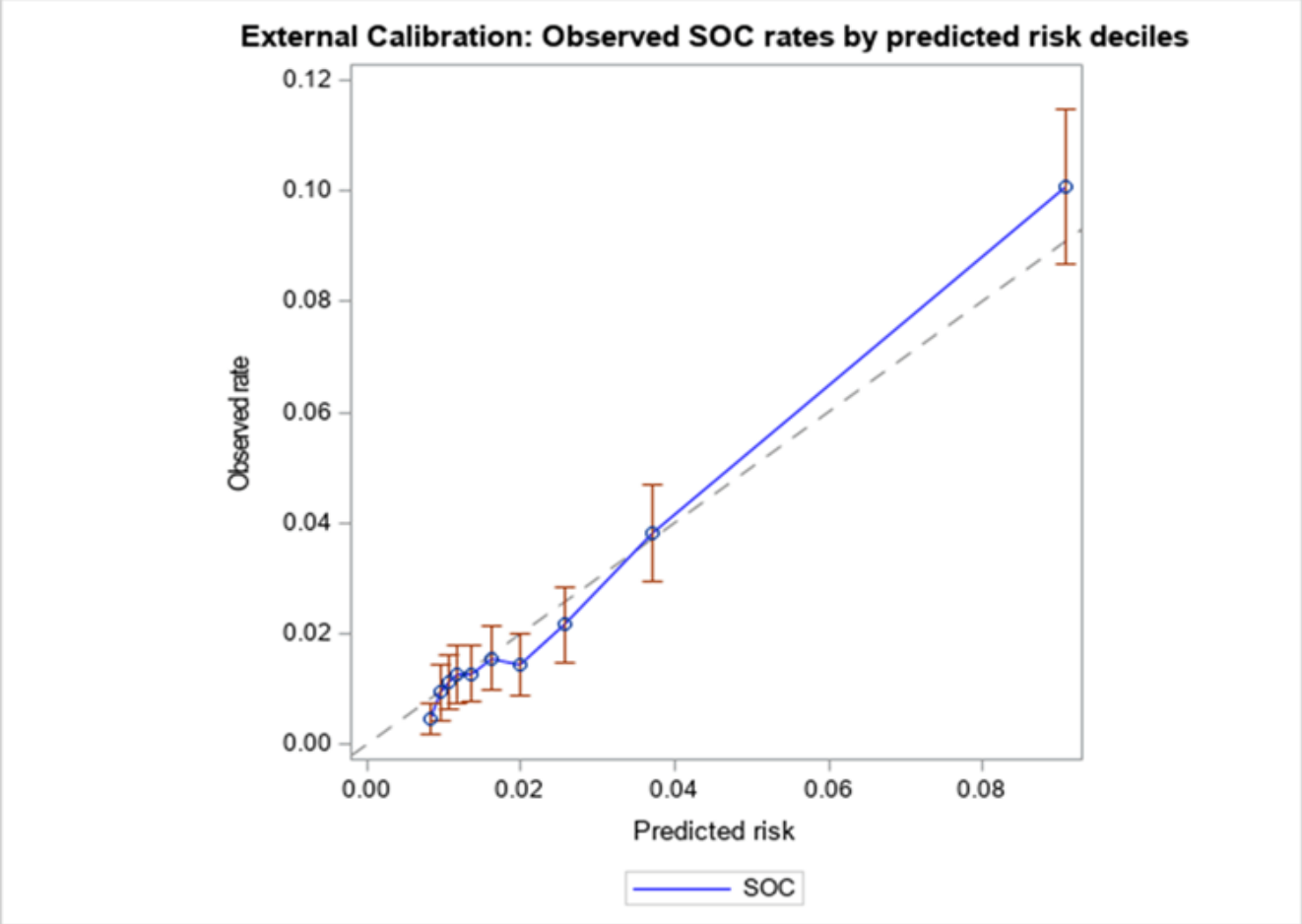


Figure 2b.29.03 Internal Calibration for Rates of Severe Obstetric Complications excluding Blood Transfusion Only Encounters

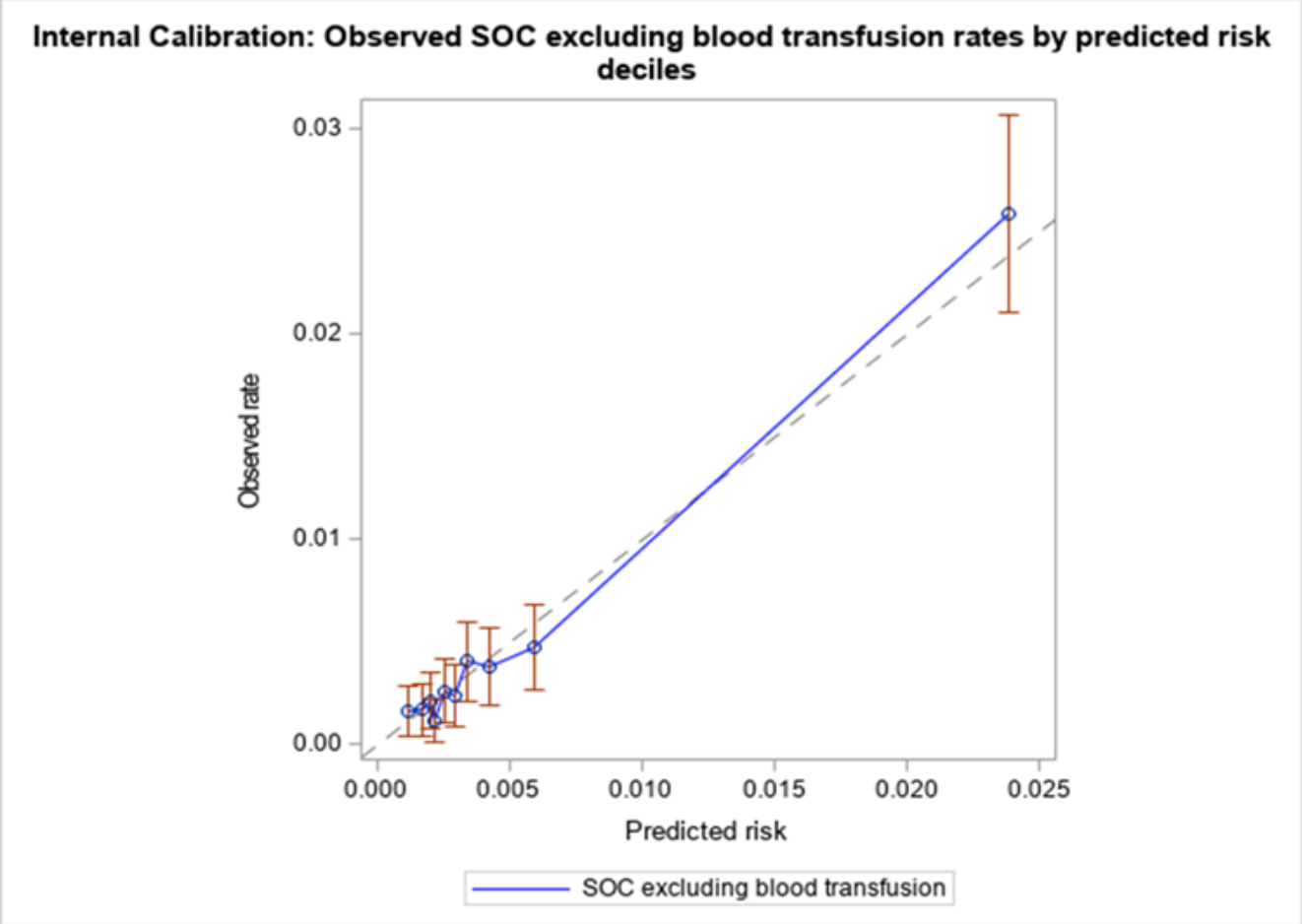
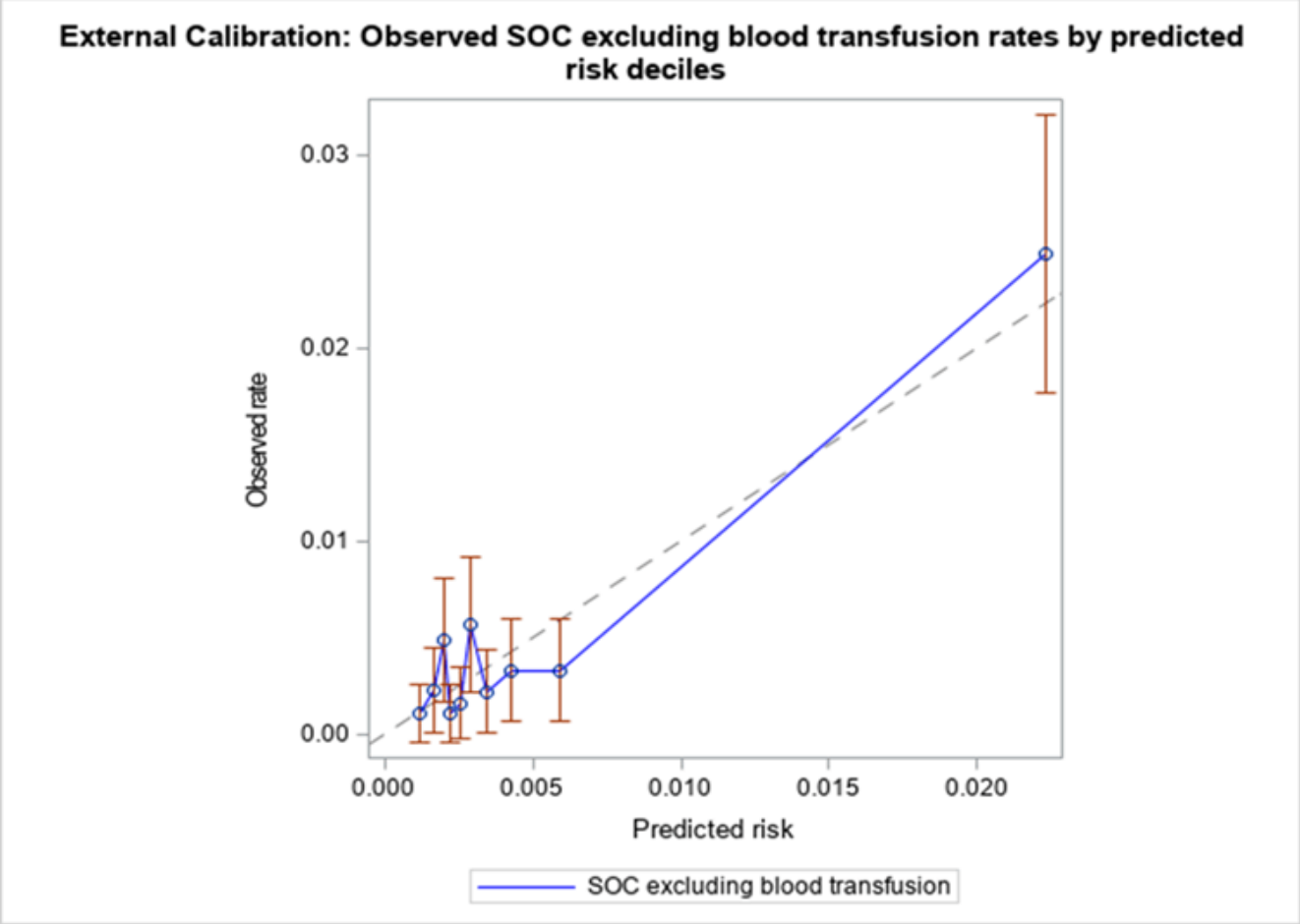


Figure 2b.29.04 External Calibration for Rates of Severe Obstetric Complications excluding Blood Transfusion Only Encounters



[Response Ends]

2b.30. Provide the results of the risk stratification analysis.

[Response Begins]
Work in progress for race/ethnicity analyses, no results as of submission.
[Response Ends]

2b.31. Provide your interpretation of the results, in terms of demonstrating adequacy of controlling for differences in patient characteristics (i.e., case mix).

In other words, what do the results mean and what are the norms for the test conducted?

[Response Begins]
The calculated C-statistics of 0.74 and 0.75 for the risk model for any severe obstetric complications (development and validation datasets), and 0.77 and 0.73 for the severe obstetric complications excluding blood transfusion-only cases measure (development and validation datasets), indicate good model discrimination.
Risk models for both the development and validation datasets show a reasonable range between the lowest decile and highest decile of predicted ability, given the low prevalence of the outcome. Overall, these diagnostic results demonstrate the risk-adjustment model adequately controls for differences in patient characteristics.
The calibration values which are consistently close to 0 at one end and close to 1 at the other end indicate good calibration of the model. If the γ0 in the model performance using validation data is substantially far from zero and the γ1 is substantially far from 1, there is potential evidence of over-fitting.
The two predictive models we created and tested had area under the ROC curve of 0.74 and 0.77 for any severe obstetric complications and severe obstetric complications excluding blood transfusion only cases, respectively. This moderate level of predictive ability demonstrates that we have identified patient characteristics related to severe obstetric

complications, and therefore controlling for these covariates in measure calculations should adequately control for differences in patient characteristics across hospitals. For risk variable selection, we favored a clinical/theoretical approach over a data-driven approach when possible, and we aimed to pre-specify all available patient characteristics present on admission that could be causally related to the outcome of severe obstetric complications.

2b.31.01 ROC Curve for Model of Any Severe Obstetric Complication

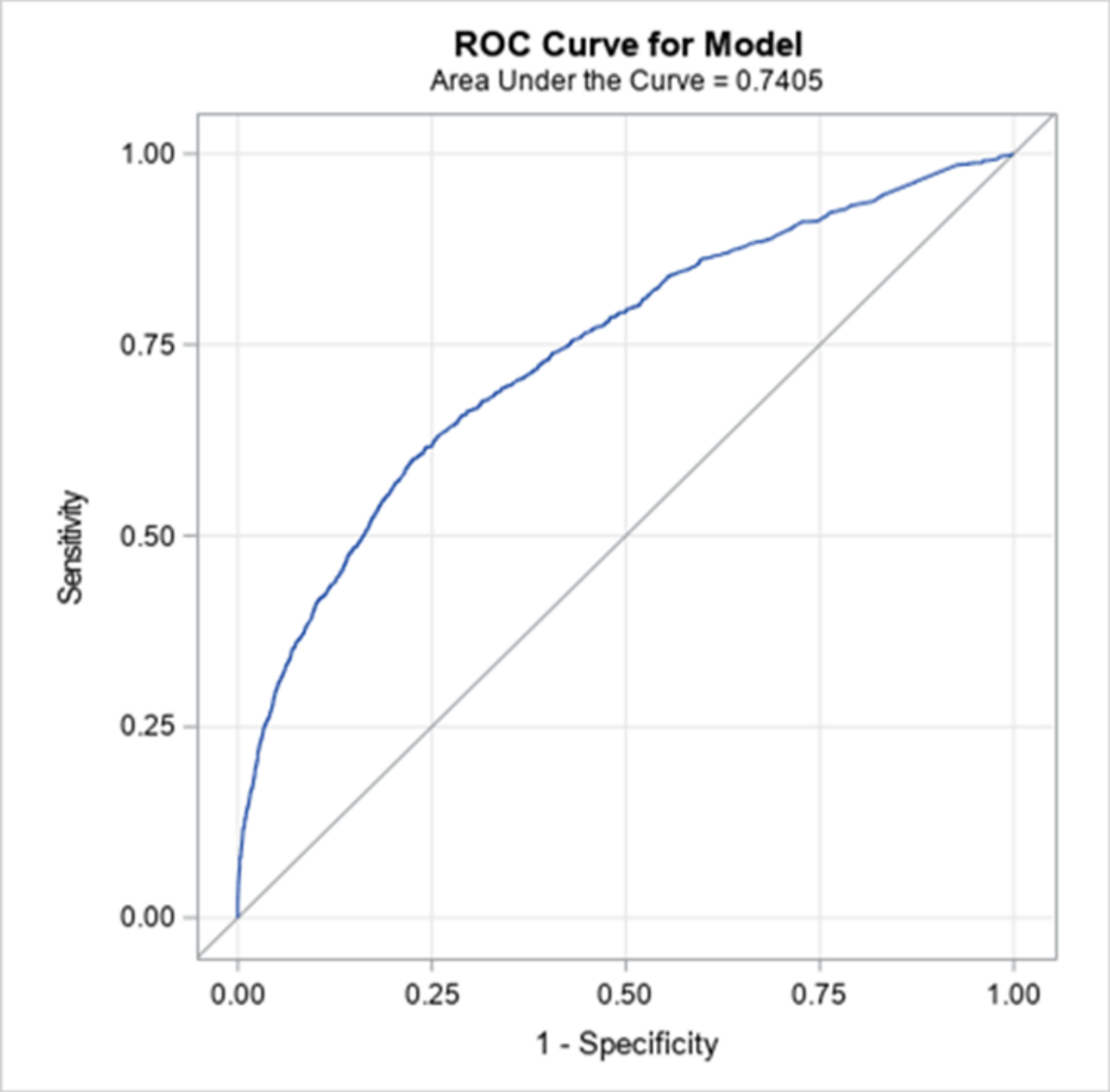
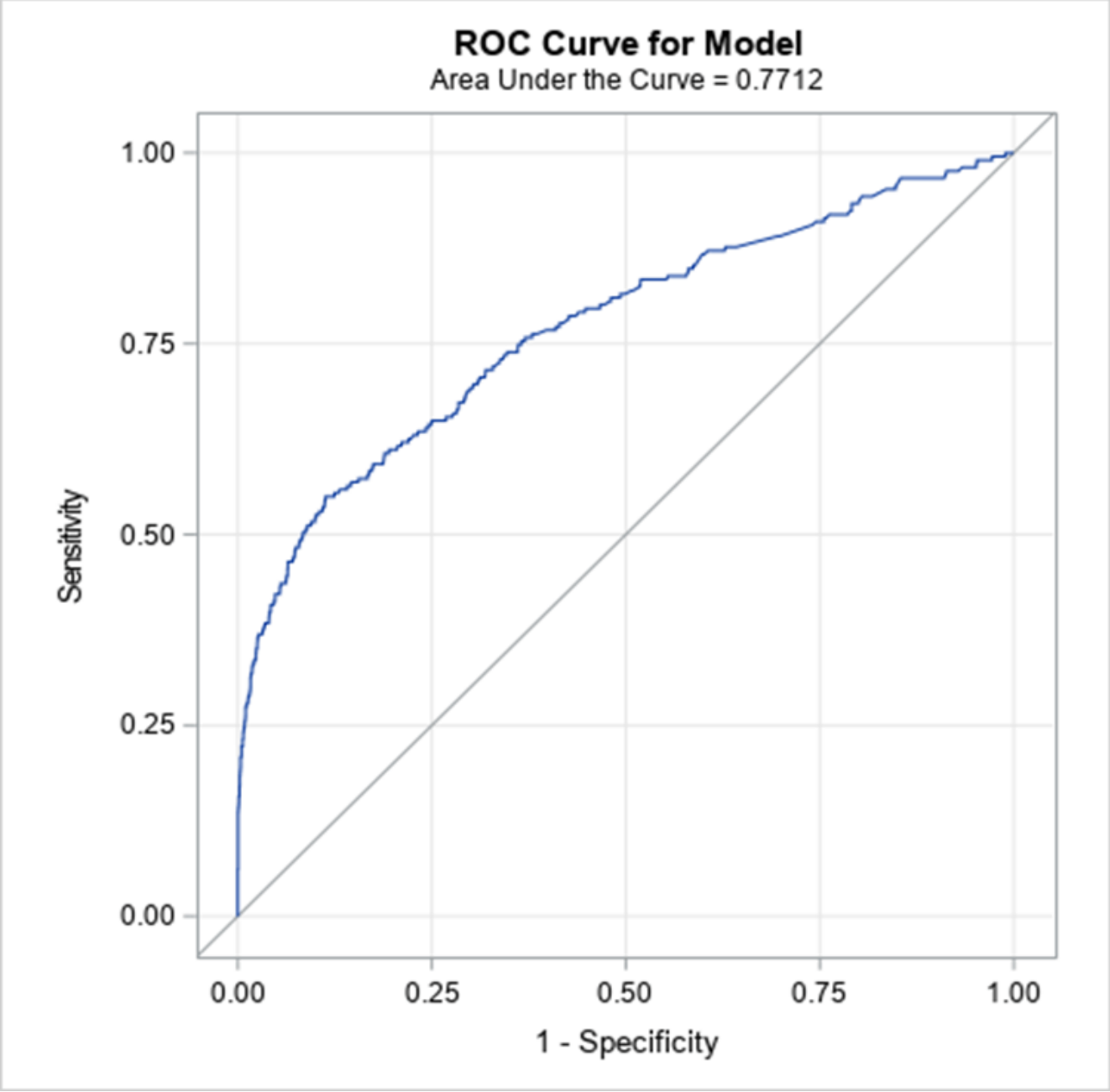


Figure 2b.31.02 ROC Curve for Model of Any Severe Obstetric Complication excluding Blood Transfusion Only Encounters



[Response Ends]

2b.32. Describe any additional testing conducted to justify the risk adjustment approach used in specifying the measure.

Not required but would provide additional support of adequacy of the risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed.

[Response Begins]
Model fit was also assessed using model Chi-square which shows the models are significantly better than the null models.
[Response Ends]

Criteria 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.

3.01. Check all methods below that are used to generate the data elements needed to compute the measure score.

[Response Begins]
Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score)
Coded by someone other than person obtaining original information (e.g., DRG, ICD-10 codes on claims)
[Response Ends]

3.02. Detail to what extent the specified data elements are available electronically in defined fields.

In other words, indicate whether data elements that are needed to compute the performance measure score are in defined, computer-readable fields.
[Response Begins]
ALL data elements are in defined fields in a combination of electronic sources
[Response Ends]

3.03. 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 data elements not from electronic sources.

[Response Begins]
Not applicable.
[Response Ends]

3.05. Complete and attach the NQF Feasibility Score Card.

[Response Begins]
See attachment.
[Response Ends]

Attachment: 3687e_PC07_nqf_ecqm_feasibility_final_scorecard_October_MUC_Submission_(1).xlsx

3.06. 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.

[Response Begins]
For feasibility testing, virtual EHR Walkthroughs were conducted with nine healthcare sites consisting of 27 individual hospitals, representing three different EHR systems. Feasibility testing included assessment of clinical and documentation workflows compared to measure intent, assessment of data element availability and accuracy, and assessment of use of data standards. Subsequent to the fourth EHR walkthrough, The Joint Commission staff determined several of the pilot sites were unable to accurately capture 2 main data elements: the timestamp for the procedure performed and the laboratory test result of the Pa02/Fi02 ratio. The Joint Commission staff proposed to address these feasibility challenges by revising the draft specifications to better align with clinical intent and decrease burden for a lab result that was not commonly calculated in the EHR. Consequently, feasibility scores based on the revised specifications increased to 98%.

Table 3.06.01 Overall Feasibility Rates

PILOT SITES	FEASIBLITY RATE 1 Initial	FEASIBILITY RATE 2 Revised
Pilot Site 1	97%	97%
Pilot Site 2	87%	94%

PILOT SITES	FEASIBLITY RATE 1 Initial	FEASIBILITY RATE 2 Revised
Pilot Site 3	97%	100%
Pilot Site 4	97%	97%
Pilot Site 5	96%	98%
Pilot Site 6	91%	100%
Pilot Site 7	97%	100%
Pilot Site 8	97%	100%
Pilot Site 9	90%	99%
Overall	95%	98%

Feasibility Rate 1: reflects the rate inclusive of the timestamp for the procedure performed and the laboratory test result of the Pa02/Fi02 ratio.

Feasibility Rate 2: reflects the rate with the revised specifications, using date only for procedures performed (no timestamp) and laboratory test results of PaO2.

Table 3.06.02 Feasibility Rates by Domain

PILOT SITES	DATA AVAILABILITY	DATA ACCURACY	DATA STANDARDS	WORKFLOW
Pilot Site 1	100%	100%	87%	100%
Pilot Site 2	94%	94%	94%	94%
Pilot Site 3	100%	100%	100%	100%
Pilot Site 4	96%	99%	96%	99%
Pilot Site 5	100%	100%	94%	99%
Pilot Site 6	100%	100%	100%	100%
Pilot Site 7	100%	100%	100%	100%
Pilot Site 8	100%	100%	100%	100%
Pilot Site 9	100%	100%	96%	100%
Overall	99%	99%	96%	99%

This table shows the feasibility rates by domain reflecting the revised specifications.

Based on an overall feasibility score of 98%, ePC07 data elements were found to be highly feasible. Validity testing showed an overall data element agreement rate of 90.4% and an overall measure outcome agreement rate of 91.2% (see Tables 2b.03.03 and 2b.03.04 for more details of findings).

Specific feedback obtained from feasibility testing are listed below. Other findings were site specific and changes to the measure specifications were not deemed necessary.

- Sites were unable to accurately capture 2 main data elements: the timestamp for the procedure performed and the laboratory test result of the Pa02/Fi02 ratio. The PaO2/FiO2 ratio was replaced with the PaO2 lab value. One site erroneously mapped the baby's cord blood PaO2 level instead of the mother's PaO2 level.
- Platelet count alone was not specific enough to identify a severe obstetric complication. During reliability visits, we saw that most cases included codes from the risk adjustment value set for anemia or bleeding disorders and did not require additional treatment or longer length of stay. Most of these cases had platelet levels that were lower on admission and fluctuated above and below 100.
- In the original version of the logic, the denominator exclusion was stated as inpatient hospitalizations for patients with trauma complicating childbirth diagnoses. Pilot testing revealed no cases where the trauma was an indication for delivery or had an impact on care. The trauma code is used too broadly in the field and does not represent the clinical intent for exclusion.
- It is common practice for hospitals to admit laboring patients to an OB Triage status until true labor is confirmed. This is an outpatient status where critical elements of care are performed. If the patient is ultimately admitted, the care rendered in the outpatient setting will not be evaluated if the logic only qualifies on the inpatient encounter.
- POA codes are not consistently assigned to SNOMED codes.

[Response Ends]

Consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

3.07. Detail 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),

Attach the fee schedule here, if applicable.

[Response Begins]
Not applicable
[Response Ends]

Criteria 4: Use and Usability

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.

Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making.
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 demonstrating performance improvement.

4a.01. Check all current uses. For each current use checked, please provide:

Name of program and sponsor

URL

Purpose

Geographic area and number and percentage of accountable entities and patients included

Level of measurement and setting

[Response Begins]
Regulatory and Accreditation Programs
[Regulatory and Accreditation Programs Please Explain]

- **Name of program and sponsor:** ORYX Performance Measure Reporting: Hospital Accreditation Program (HAP) and Critical Access Hospital Accreditation (CAH) Program, The Joint Commission
- **URL:** https://www.jointcommission.org/measurement/reporting/accreditation-oryx/
- **Purpose:** An accreditation program that recognizes hospitals that meet standard requirements to provide safe and effective patient care.
- **Geographic area and number and percentage of accountable entities and patients included:** The Joint Commission accredits 63% of hospitals, 81% of beds; participating hospitals with maternity services includes >2500 US hospitals Nationwide. First year in production. No production data available.
- **Level of measurement and setting:** Outcome measure inpatient delivery hospitalization, all TJC participating hospitals with maternity services

Quality Improvement with Benchmarking (external benchmarking to multiple organizations)
[Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Please Explain]

- **Name of program and sponsor:** ORYX Performance Measure Reporting: Hospital Accreditation Program (HAP) and Critical Access Hospital Accreditation (CAH) Program, The Joint Commission
- **URL:** https://www.jointcommission.org/measurement/reporting/accreditation-oryx/
- **Purpose:** An accreditation program that recognizes hospitals that meet standard requirements to provide safe and effective patient care. The data submitted to The Joint Commission is analyzed for trends and benchmarks.
- **Geographic area and number and percentage of accountable entities and patients included:** The Joint Commission accredits 63% of hospitals, 81% of beds; participating hospitals with maternity services includes >2500 US hospitals Nationwide. First year in production. No production data available.
- **Level of measurement and setting:** Outcome measure inpatient delivery hospitalization, all TJC participating hospitals with maternity services

Quality Improvement (Internal to the specific organization)
[Quality Improvement (Internal to the specific organization) Please Explain]

- **Name of program and sponsor:** ORYX Performance Measure Reporting: Hospital Accreditation Program (HAP) and Critical Access Hospital Accreditation (CAH) Program, The Joint Commission

- **URL:** https://www.jointcommission.org/measurement/reporting/accreditation-oryx/
- **Purpose:** An accreditation program that recognizes hospitals that meet standard requirements to provide safe and effective patient care. The data submitted to The Joint Commission is analyzed for trends and benchmarks and provided to the organizations for internal quality improvement purposes.
- **Geographic area and number and percentage of accountable entities and patients included:** The Joint Commission accredits 63% of hospitals, 81% of beds; participating hospitals with maternity services includes >2500 US hospitals Nationwide. First year in production. No production data available.
- **Level of measurement and setting:** Outcome measure inpatient delivery hospitalization, all TJC participating hospitals with maternity services

[Response Ends]

4a.02. Check all planned uses.

[Response Begins]
Public reporting
Measure Currently in Use
[Response Ends]

4a.03. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing), explain why the measure is not in use.

For example, do policies or actions of the developer/steward or accountable entities restrict access to performance results or block implementation?

[Response Begins]
[Response Ends]

4a.04. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes: used in any accountability application within 3 years, and publicly reported within 6 years of initial endorsement.

A 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.

[Response Begins]
[Response Ends]

4a.05. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

Detail 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.

[Response Begins]
For reference, each health system will be referred to as a 'pilot site' and 'hospital' will refer to the individual hospitals within the health system. A total of 10 pilot sites consisting of 28 hospitals were included in the pilot project. For feasibility testing, 9 pilot sites with a total of 27 hospitals were included for analysis. After feasibility testing, 1 pilot site representing 2 hospitals withdrew from the project and one additional hospital was added. Therefore, data was collected from 9 pilot sites representing 26 hospitals. Reliability and validity testing was completed on 6 sites representing 15 hospitals.
After the pilot testing concluded and final results were analyzed, a pilot summary report was created and shared with each pilot site via email. Contents of the summary report were presented in a clear manner, with the purpose of each

testing modality explained along with information on how to interpret the results of statistical testing. The pilot summary included general measure information, feasibility, reliability and validity testing, risk model, and performance results. Each pilot site received their own individual site measure results and analysis along with the aggregate pilot summary report. Prior to the pilot testing, Joint Commission staff provided virtual information sessions reviewing measure specifications, pilot testing overview and an EHR walkthrough session. Q&A opportunities were provided to the sites. Joint Commission staff also offered assistance to the pilot sites for any questions they had regarding the pilot summary reports.

[Response Ends]

4a.06. Describe the process for providing measure results, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

[Response Begins]
Upon completion of testing, a live national webinar was held on March 8, 2022, to introduce the ePC07 measure including a detailed explanation of the specifications. The webinar included an opportunity for audience members to ask questions.
Severe Obstetric Complications is a new measure, and our implementation plan includes continuous customer engagement. The Joint Commission developed dashboards as part of the ongoing continuous customer engagement project. The dashboard report—posted in the Resources and Tools section of an accredited hospital's secure Joint Commission Connect® extranet site—is representative of each organization's relative performance on each of the selected measures. For each measure, the dashboard shows that organization's performance compared to national, state, and Joint Commission–accredited organization averages. The dashboard is not a scorable element on the survey, but rather, a tool to facilitate discussion about ongoing quality improvement work. For example, surveyors may ask an organization how it addresses the subset of performance measures in the report and what action(s) the organization is taking to improve processes. In addition, the Joint Commission analyzes aggregate performance of each measure and identifies the measures for which the greatest opportunities for improvement exist among accredited hospitals. Based on those findings, an educational webinar series that address the high-opportunity topics is developed. All accredited hospitals have access to the educational webinar series. Organizations with high opportunity for improvement are particularly encouraged to participate.

[Response Ends]

4a.07. Summarize the feedback on measure performance and implementation from the measured entities and others. Describe how feedback was obtained.

[Response Begins]
Since ePC07 was recently published in January of 2022, we do not have measure performance data as of yet. However, we were able to obtain feedback during the pilot testing of this measure. See section 4a.05 for details on pilot test sites. Feedback was also obtained through Technical Expert Panel meetings and surveys, Patient Workgroup meetings and surveys, and public comment.
The Joint Commission plans to use an automated feedback system currently used for feedback on other measures. Access is available to the measured entities and the vendors contracted by measured entities. The measure leads from the clinical team and the eCQM team are responsible for each individual measure set. The system is monitored daily, and responses are typically provided within 8 business hours.

[Response Ends]

4a.08. Summarize the feedback obtained from those being measured.

[Response Begins]
During pilot site recruitment and engagement, feedback received from hospitals indicated that leadership teams were interested in the measure, and development of a Severe Obstetric Complications measure was vital and of great value. One hospital was planning on adding the ePC-07 metric to their annual dashboard for future use.

Feedback Obtained During Public Comment:

- The Call for Public Comment ran from November 19, 2021, to December 18, 2021.
- The measure developer solicited public comments by email notification to CMS listserv groups, emails to relevant stakeholders and stakeholder organizations, and posting on the CMS Public Comment website. We received eighteen responses on this topic.
- Some highlights of the public comment are that commenters provided support for:
  - focusing measurement on addressing severe maternal morbidity and improving maternal health outcomes.
  - the usefulness of this measure in assessing and improving the quality of care for patients.
  - publicly reporting both an overall rate of severe obstetric complications and a rate of severe obstetric complications excluding blood transfusion-only cases.

[Response Ends]

4a.09. Summarize the feedback obtained from other users.

[Response Begins]

- The face validity assessment demonstrated that the Technical Expert Panel members believe that this eCQM is an important health outcome to measure because there is room for improvement, it will produce reliable and valid rates, and hospitals can use the results for performance improvement. While there are some concerns with the feasibility of implementation and whether this measure is a critical component of defining and comparing the quality of obstetric care between hospitals, the majority of the responses from the TEP either agreed or strongly agreed with the ability of this measure to improve patient outcomes. See Section 2b.03 for further details on face validity.
- As described in 1a.02, the Patient Working Group members strongly believe this eCQM is an important health outcome to measure because there is room for improvement and strongly/moderately agree that this measure is a critical component of defining and comparing the quality of obstetric care between hospitals. See Section 2b.03 for further details on face validity.

[Response Ends]

4a.10. Describe how the feedback described has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

[Response Begins]

- As mentioned in 3.06, pilot sites were unable to accurately capture 2 main data elements: the timestamp for the procedure performed and the laboratory test result of the Pa02/Fi02 ratio. The Joint Commission addressed these feasibility challenges by revising the draft specifications to better align with clinical intent and decrease burden for a lab result not commonly calculated in the EHR. The PaO2/FiO2 ratio was replaced with the PaO2 lab value which was removed from the final specifications as it was found to be a low volume test and mapping was burdensome.
- Platelet count < 100 10*3/uL was removed from the numerator (see 4a.08 for reason).
- Trauma was removed from the denominator exclusion logic (see 4a.08 for reason).
- A denominator exclusion for COVID plus respiratory conditions was added post pilot due to the growing evidence of perinatal complications in women who have COVID-19 infection with respiratory conditions.
- To account for care rendered in an outpatient setting, the logic evaluates any care rendered in the Emergency Department, observation or OB Triage areas within one hour of inpatient admission.

Since pilot testing revealed that POA codes are not consistently assigned to SNOMED codes, SNOMED codes were removed from most numerator and risk variable value sets. It is important that this measure discerns that a severe obstetric complication was not present on admission (POA) and that any condition used for risk adjustment was POA. POA code assignment for ICD10 codes is thoroughly adopted and implemented by healthcare organizations. We recognize the importance and value of SNOMED codes and have therefore developed draft value sets for SNOMED codes for use in future versions of the measure specifically in the numerator and risk variables. We will continue to investigate the feasibility of implementing SNOMED codes with POA codes to allow for use in the measure logic and ensure clinical intent.

[Response Ends]

4b.01. You may refer to data provided in Importance to Measure and Report: Gap in Care/Disparities, 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, provide an explanation. 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.

[Response Begins]
This is a de novo eCQM intended to measure inpatient acute care hospital quality and performance related to severe obstetric complications and death during the delivery hospitalization. The measure is intended to be used alongside the suite of existing perinatal process of care quality measures and existing quality improvement efforts focused on reducing maternal morbidity and mortality.
Although there are limited measures to assess variability among hospitals, rates in the United States are higher than all other developed countries, presenting an opportunity for improvement. Using the CDC definition of SMM, the US median rate was 1.4% and the highest hospital rate was 12.2%.29 USA Today's database of childbirth complication rates at maternity hospitals, with data from 1,027 hospitals in 13 states from 2014-2017, showed marked variation in median rates of childbirth complications; this variability may reflect similar trends for maternal complications.1,3
Maternal morbidity has garnered a lot of national attention, with a broad range of SMM events and outcomes that can be examined, many of which are closely associated with mortality.2,3 Several initiatives have shown promise in reducing maternal morbidity events. For example, since the inception of the California Maternal Quality Care Collaborative (CMQCC), focused on metrics and toolkits to improve maternal outcomes, the maternal mortality rate in California declined by 55% between 2006 and 2013.4 The CMQCC obstetric hemorrhage collaborative resulted in a 20.8% reduction in SMM in California hospitals compared with the 1.2% reduction in SMM among nonparticipating hospitals.3 The state of California has established a successful framework for assessing and improving quality of maternal care, and outcomes suggest great potential for nationally reducing maternal care complications.
State and national initiatives to measure, track, and reduce maternal morbidity and mortality have produced encouraging results. The Severe Obstetric Complications eCQM could expand these improvements in care, outcomes, and cost savings at a national level. The eCQM will provide hospitals with benchmarking and actionable data to inform their quality improvement efforts; the use of EHR data will provide them with the potential to repurpose the data and measure logic for internal quality control using real-time feedback to further mitigate harm to mothers. Additionally, the eCQM can provide information that allows patients to compare hospitals' performance to aid in their decision making when choosing care.
Additional information can be found in 1a.03.
1. Deadly Deliveries: Childbirth complication rates at maternity hospitals. https://www.usatoday.com/maternal-mortality-harm-hospital-database/.
2. National Quality Forum. Maternal Morbidity and Mortality Environmental Scan. 2020.
3. Main EK. Reducing maternal mortality and severe maternal morbidity through state-based quality improvement initiatives. Clinical obstetrics and gynecology. 2018;61(2):319-331.
4. California Maternal Quality Care Collaborative (CMQCC). Who We Are. https://www.cmqcc.org/who-we-are, 2020.

[Response Ends]

4b.02. Explain any unexpected findings (positive or negative) during implementation of this measure, including unintended impacts on patients.

[Response Begins]
The measure specifications were posted January 28, 2022, for optional use in the Joint Commission ORYX Performance Measure Reporting Requirements: Hospital Accreditation Program (HAP) and Critical Access Hospital Accreditation (CAH) Program. No implementation findings at this time. Data will be submitted to The Joint Commission in 2023 for optional year 2022.
Potential unintended consequences: Measuring obstetric complication outcomes based on EHR data may cause a shift in a hospital's resources to support EHR data extraction and reporting, and away from other functions. Also, although the measure numerator definition is broad, hospitals may potentially focus on complications captured in the measure, while dismissing other complications not currently measured but that are important, as well.

[Response Ends]

4b.03. Explain any unexpected benefits realized from implementation of this measure.

[Response Begins]
The measure specifications were posted January 28, 2022, for optional use in the Joint Commission ORYX Performance Measure Reporting Requirements: Hospital Accreditation Program (HAP) and Critical Access Hospital Accreditation (CAH) Program. No implementation findings at this time.
[Response Ends]

Criteria 5: Related and 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.

If you are updating a maintenance measure submission for the first time in MIMS, please note that the previous related and competing data appearing in question 5.03 may need to be entered in to 5.01 and 5.02, if the measures are NQF endorsed. Please review and update questions 5.01, 5.02, and 5.03 accordingly.

5.01. Search and select all NQF-endorsed related measures (conceptually, either same measure focus or target population).

(Can search and select measures.)
[Response Begins]
[Response Ends]

5.02. Search and select all NQF-endorsed competing measures (conceptually, the measures have both the same measure focus or target population).

(Can search and select measures.)
[Response Begins]
[Response Ends]

5.03. If there are related or competing measures to this measure, but they are not NQF-endorsed, please indicate the measure title and steward.

[Response Begins]
No related or competing measures.
[Response Ends]

5.04. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s), indicate whether the measure specifications are harmonized to the extent possible.

[Response Begins]
No
[Response Ends]

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

[Response Begins]
Not applicable. No related or competing measures.
[Response Ends]

5.06. Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality). Alternatively, justify endorsing an additional measure.

Provide analyses when possible.

[Response Begins]
National evaluation of hospitals' performance on maternal morbidity and mortality is limited because there are currently no maternal morbidity or obstetric complications outcome measures in national reporting programs. Current quality measures related to pregnancy and maternal health proposed for or in public reporting programs are largely process

measures (e.g., Maternity Care: Post-partum Follow Up and Care Coordination) and outcome measures related to delivery type (e.g., PC-01 Elective Delivery).

There are numerous state agencies, private and/or non-profit organizations, and collaboratives that have spearheaded maternal health and quality improvement initiatives. For instance, the Alliance for Innovation in Maternal Health (AIM) developed evidence-based patient safety bundles to address leading causes of SMM, like obstetric hemorrhage and hypertension. The CDC Perinatal Collaboratives also support various state-based efforts to promote high quality maternal care. The CMQCC created the Maternal Data Center (MDC) for hospitals with Labor and Delivery units in California, Oregon, and Washington. The MDC is an online tool that receives patient discharge data on maternity care services, linking these data to birth certificate or clinical data, and feeding back to clinicians' perinatal performance data for supporting quality improvement.1 The MDC allows hospital performance regional and statewide comparisons. Overall, such quality metrics do not currently cater to a national population because there is extensive variation and timing delays in the widespread adoption and implementation of safety protocols in obstetric care across states.2,3 Moreover, data examining the nationwide implementation of these resources are not widely available.2,4 Therefore, the development of a obstetric complications outcome measure addresses a national measurement gap that can build on learnings from existing maternal health initiatives and measures.

1. California Maternal Quality Care Collaborative (CMQCC). Maternal Data Center. https://www.cmqcc.org/maternal-data-center, 2020.
2. Main EK. Reducing maternal mortality and severe maternal morbidity through state-based quality improvement initiatives. Clinical obstetrics and gynecology. 2018;61(2):319-331.
3. Lenfant C. Clinical research to clinical practice—lost in translation? New England Journal of Medicine. 2003;349(9):868-874.
4. Maher-Griffiths C. Maternal Quality Outcomes and Cost. Critical Care Nursing Clinics. 2019;31(2):177-193.

[Response Ends]