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 #: 0471e

Corresponding Measures:

Measure Title: ePC-02 Cesarean Birth

Measure Steward: The Joint Commission

Brief Description of Measure: This measure assesses the number of nulliparous women with a term, singleton baby in a vertex position delivered by cesarean birth.

Developer Rationale: The removal of any pressure to not perform a cesarean birth has led to a skyrocketing of hospital, state and national cesarean birth (CB) rates. Some hospitals' CB rates were over 50%. Hospitals with CB rates at 15-20% have infant outcomes that are just as good and better maternal outcomes (Symum et al., 2021). There is no data that higher rates improve any outcomes, yet the CB rates continue to rise. This measure seeks to focus attention on the most variable portion of the CB epidemic, the term labor CB in nulliparous women. This population segment accounts for the large majority of the variable portion of the CB rate and is the area most affected by subjectivity.

As compared to other CB measures, what is different about the nulliparous, term, singleton, vertex (NTSV) CB rate (Primary CB in first births with term singleton pregnancies in head down position) is that there are clear cut quality improvement activities that can be done to address the differences. Main et al. (2012) found that over 60% of the variation among hospitals can be attributed to first birth labor induction rates and first birth early labor admission rates. The results showed if labor was forced when the cervix was not ready the outcomes were poorer. Rosenstein et al. (2021) also showed that labor and delivery guidelines can make a difference in labor outcomes. Many authors have shown that physician factors, rather than patient characteristics or obstetric diagnoses are the major driver for the difference in rates within a hospital (Berkowitz, et al.,1989; Goyert et al., 1989; Luthy et al., 2003, Symum et al., 2021). The dramatic variation in cesarean rates seen in all populations studied is striking. (Cesarean rates varied tenfold in US hospitals nationwide across hospitals, from 7.1 % to 69.9 % and there was a 15-fold variation among low-risk women, from 2.4% to 36.5% (Kozhimannil et al., 2013).

A reduction in the number of nulliparous patients with live term singleton newborns in vertex position (NTSV) delivering by cesarean birth will result in increased patient safety, a substantial decrease in maternal and neonatal morbidity and substantial savings in health care costs. Successful quality improvement efforts incorporate audit and feedback strategies combined with provider and nurse education, guidelines and peer review.

The measure will assist health care organizations (HCOs) to track nulliparous patients with live term singleton newborns in vertex position delivering by cesarean birth to reduce the occurrence. Nulliparous women have 4-6 times the cesarean birth rate than multiparous women thus the NTSV population is the largest driver of primary cesarean birth rate (Sakala et al. 2020). NTSV has a large variation among facilities, thus identifying an important population on which to focus quality improvement efforts.

In addition, a reduction in primary cesarean births will reduce the number of women having repeat cesarean births (almost 90 of mothers who have a primary cesarean birth will have subsequent cesarean birth (CDC, 2020)). Thus, improvement in the rates of cesarean birth for the first birth will reduce the morbidity of all future births and avoid all the controversies with trial of labor after cesarean/elective repeat cesareans.

Numerator Statement: Inpatient hospitalizations for patients who deliver by cesarean section.

Denominator Statement: Inpatient hospitalizations for nulliparous patients delivered of a live term singleton newborn >= 37 weeks gestation.

Denominator Exclusions: Inpatient hospitalizations for patients with abnormal presentation or placenta previa during the encounter.

Measure Type: Outcome

Data Source: Electronic Health Records, Electronic Health Data

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 outcome electronic clinical quality measure (eCQM) at the facility level of analysis that assesses the number of nulliparous women with a term, singleton baby in a vertex position delivered by cesarean birth.
- The developer provides a logic model that depicts the healthcare organization identifying NTSV patients, which leads to the healthcare organization supporting vaginal delivery methods for these patients, which leads to fewer cesareans for first-time delivering persons, which leads to reduced obstetric and neonatal morbidities.

Summary:

- To demonstrate the value of the measure to patients, the developer cites research showing that roughly 30 percent of patients who had a cesarean delivery actively sought out information on cesarean rates at their hospital.

- To demonstrate the relationship between the outcome and provision of care, the developer cites the American College of Obstetricians and Gynecologists (ACOG) recommendation of reduction of cesarean rates in the NTSV population and methods for reduction (increasing recommended hours of "pushing" for NTSV patients, increased training in use of forceps or manual rotation/aversion) and the benefits for reducing the repeat cesarean rate.
- The developer also cites research to demonstrate the effect of a large-scale improvement collaborative to reduce NTSV cesarean delivery rates in California hospitals (total annual delivery volume of 119,000 birthing individuals). After implementation of the collaborative, cesarean rates decreased, and maternal and neonatal morbidities did not increase.

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 (Box 1) -> the relationship between the measured/patient reported health outcome and at least one healthcare action is demonstrated by empirical data -> 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.

- The developer has set 'a rate greater than 30 percent' as the threshold for public reporting of this measure as this is approximately two standard deviations from the mean two-year rate.
- Gap data are summarized at the hospital level for 2020 discharges. Due to a small number of hospitals (n=15) participating in the pilot study, the developer reported five number statistical summaries instead of deciles.
    - Mean - 27.5 percent (standard deviation (SD): 20.0 percent)
    - Maximum - 71.8 percent
    - Minimum - 0 percent
    - 25th Percentile - 19.5 percent
    - 50th Percentile - 23.3 percent
    - 75th Percentile - 28.9 percent
- The developer also notes that in 2020, the rate was at 27.5 percent nationally.
- The 2030 Healthy People goal for cesarean rates is 23.6 percent.

Disparities

- The developer reports measure rates by age, ethnicity, race, and payer:
    - By age, rates were highest in the 40+ group at 42.9 percent (n=28), followed by 35-40 at 33.7 percent (n=101). The range was 20.2 percent for under 20 years of age (n=94) to 42.9 percent for 40+.
    - By ethnicity, Hispanic or Latino had rates of 35.8 percent (n=296), not Hispanic or Latino of 23.9 percent (n=695).

- By race, and among those categories with a larger n, Black or African American had rates of 32.4 percent (n=105). Asian had rates of 18.9 percent (n=74). White women had a rate of 30.8 percent (n=577).
- By payer, commercial coverage had rates of 23.5 percent (n=600) while Medicaid/Medicare was at 20.3 percent (n=237).

- The developer also cites a retrospective cohort study showing that all race and ethnic categories had higher odds of cesarean delivery compared to White women.
  - Compared to White women, Black women had greater odds of fetal intolerance as an indication, while Hispanic and Asian women had greater odds of failure to progress.
  - Disparities in cesarean delivery rates were not explained by maternal, neonate, or facility factors.

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

- Rates of c-sections remain high; first time births is a real opportunity for QI in this area (especially given limited offerings for VBACs)
- New electronic measure. Good evidence to support measure focus.
- Strong evidence

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

- Saw variation across sites and across different racial/ethnic groups
- Clear performance gaps exist on many levels.
- Large performance gap across the nation with racial and ethnicity opportunities for improvement.

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; Eric Weinhandl (Combined Methods Panel Review)

- The SMP Did Not Reach Consensus on Reliability with a score of: H-0; M-4; L-3; I-2
- The SMP Did Not Reach Consensus on Validity with a score of: H-0; M-5; L-2; 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 (QDM, HQMF, and 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 clear and precise.
- eCQMs was specified using the latest industry accepted eCQM technical specifications: health quality measure format (HQMF), Quality Data Model (QDM), Clinical Quality Language (CQL), and value sets vetted through the National Library of Medicine's Value Set Authority Center (VSAC).

Reliability Testing:

- Reliability testing conducted at the Patient/Encounter Level:
    - The developer utilized patient/encounter level validity testing (details below) to demonstrate patient/encounter level reliability.

SMP Summary:

- The SMP recognized that the developer utilized patient/encounter level validity testing to demonstrate patient/encounter level reliability and briefly discussed reliability concerns.
- The developer clarified for the SMP that difficulties with reporting data for one test site were resolved by placing data elements into discrete fields.
- The SMP re-voted on reliability following the discussion and ultimately did not reach consensus on this criterion.

Questions for the Committee regarding reliability:

- *Do you have any concerns that the measure cannot be consistently implemented (i.e., are measure specifications adequate)?*

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

- SMP did not reach consensus

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 utilized data from two sites representing seven hospitals. A representative sample of the electronically submitted inpatient encounters was selected for re-abstraction.
    - Specificity was high for both sites. Specificity was 96.5 percent for Site 1 and 100 percent for Site 2, and 97.7 percent overall.
    - Sensitivity was high for Site 1, but low for Site 2. The sensitivity was 87.5 percent for Site 1, 0 percent for Site 2 and 73.7 percent overall.
        - The developer explained Site 2's low sensitivity by noting that no numerator events were initially identified in submitted data. Cases did not qualify for the initial population due to missing time of delivery or incorrect noting of gravida/para/term/preterm. Site 2 used a standalone OB documentation system that does not interface completely with Meditech. OB documentation is present in Meditech in non-discrete fields in a .pdf format.
        - The developer notes that they have put a mitigation plan into place for Site 2 that has eliminated this issue.
- The Feasibility Scorecard indicated that the following data elements have issues with accuracy:
    - Assessment, Performed: Estimated Gestational Age at Delivery, authorDatetime

Exclusions

- The measure excludes inpatient hospitalizations for patients with abnormal presentation or placenta previa during the encounter.
- In the exclusions analysis, it was found that exclusions had an appreciable impact on measure rates: without excluding these cases measure rates increase overall by 17 percent, or 4.7 percentage points. Exclusion rates ranged from 0-16 percent, indicating variability across sites.

Risk-Adjustment

- The measure is not risk adjusted or stratified.
- The developer gives the following rationale for not risk-adjusting the measure: exclusion criteria were chosen to ensure that the target population would be women with nulliparous, term, singleton, vertex (NTSV) pregnancies, who have a lower risk of maternal morbidity and mortality during a vaginal birth delivery than do women who have undergone a previous C-section. Therefore, the population of women in the denominator as a result of the exclusions, allow the measure to focus on a more homogeneous group of women where the greatest improvement opportunity exists as evidenced by the variation in rates of NTSV cesarean births indicating clinical practice patterns may affect this rate (ACOG, 2014).

Meaningful Differences

- The developer calculated a funnel plot for hospital rates of the measure in which the observed measure is plotted against a measure of its precision. Prediction limits are superimposed on this plot to identify outliers.
- Even with a small number of hospitals in the pilot testing, there was significant variation in measure rates and high outliers were identified.

Missing Data

- Missing data was quantitatively assessed as part of data element validity testing by indicating the "match" rate. A match indicates data was not missing and was accurate. The developer determined the percent of mismatches for each data element that were due to missing data.
    - There was variation in data completeness between Site 1 and 2 with results of 96.5 percent and 78.9 percent, respectively.
    - Site 1 had no mismatches due to missing data.
    - Site 2 has engaged in mitigation plans to improve upon the number of missing data elements. Because Pilot Site 2 uses a stand-alone OB documentation system, data elements are not in discrete fields. Most mismatches were in the Delivery Date/Time, Estimated Gestational Age, Gravida, Para, Preterm or Term Birth fields. Of the mismatches, 57 percent were due to missing data.

Comparability

- The measure only uses one set of specifications.

SMP Summary:

- During the SMP's preliminary analyses, concerns were raised regarding the failure to provide kappa results for the required data elements in the validity testing. In their response to SMP concerns, the measure developer provided an additional table to the SMP that included updated testing with kappa values. However, some SMP members expressed concerns with the kappa results.
- SMP members noted concerns with any outcome measure that is not risk adjusted. However, the SMP members agreed that the developer did provide a rationale for consideration and deferred the question on the appropriateness of the risk adjustment strategy to the Perinatal and Women's Health Standing Committee.
- Overall, some of the SMP members noted that the approach to testing reliability and validity was correct; however, the results raised reliability and validity concerns. Following the discussion, the SMP re-voted on validity but did not reach consensus on this criterion.

Questions for the Committee regarding validity:

- *Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?*
- *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**

- SMP did not reach consensus

Committee Pre-evaluation Comments:

2a. Reliability

- 2a1. Reliability-Specifications
    - It would be helpful if the developer can address the missing data at Site 2 and how this issue will be solved moving forward for all sites

    - Specifications are clear and should be reportable by any birthing facility. This is an important measure and facilities should be required to configure their data systems to ensure accurate reporting.
    - None except for provider willingness to change practice patterns/behaviors.
- 2a2. Reliability – Testing
    - None
    - No
    - No

2b. Validity

- No concerns
- No
- No

2b2-2b6. Potential threats to validity

- 2b2-3. Other Threats to Validity (Exclusions, Risk Adjustment)
    - No concerns; the measure denominator is a narrow population (low-risk, first time deliveries), so perhaps negates the need for risk-adjustment or stratification
    - The measure does not need to be risk-adjusted; the exclusion criteria are clear.
    - This measure is not risk-adjusted
- 2b4-7. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data)
    - It sounded like stand-alone OB documentation systems can result in higher rates of missing data; not clear how this will be addressed?
    - Missing data could constitute a threat to validity. However, I believe that birthing facilities should be required to refine their processes and configure their systems to eliminate missing data.
    - There are meaningful differences noted in practice with quality implications. Missing data could threaten the validity of the measure performance.

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.

- The developer attests that the data elements are generated or collected and used by healthcare personnel during the provision of care.
- The developer also notes that the data elements are coded by someone other than person obtaining 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.
    - Assessment, Performed: Estimated Gestational Age at Delivery, authorDatetime

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**

Committee Pre-evaluation Comments:

3. Feasibility

- All elements are generated during care delivery and 100% of the measure logic can be automated
- No concerns. Should be required.
- 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

- The developer offers opportunities for live feedback during meetings and by email.
- The developer notes feedback received during measure testing to improve several data elements:
    - Assessment, Performed: Date and time of obstetric delivery, Author Date Time has been replaced with Assessment, Performed: Date and time of obstetric delivery, relevant date/time
    - Assessment, Performed: Estimated Gestational Age at Delivery, Author Date Time rate has been replaced with Assessment, Performed: Estimated Gestational Age at Delivery, relevant date/time
    - Assessment, Performed: Estimated Gestational Age at Delivery, result

***Questions for the Committee**:*

- *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 (4b1. Improvement; 4b2. 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

- The developer notes that they have only one year of data, so cannot provide improvement results.

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

- During measure development, the developer noted issues with the interface between Meditech and OB documentation systems containing non-discrete fields. The developer worked with the hospital to create a mitigation plan for this issue.

Potential harms

- The developer does not note any potential harms of measure use.

Additional Feedback:

- This measure was reviewed by the Measure Applications Partnership (MAP) for the Hospital Inpatient Quality Reporting Program (Hospital IQR); EHR Incentive/EH/CAH programs in 2018. MAP recommended conditional support for rulemaking pending an evaluation and endorsement from NQF.
    - MAP noted the importance of eliminating early deliveries and improving maternal health outcomes and discussed high-risk conditions such as pre-eclampsia/eclampsia that would indicate a cesarean birth, and the implications of the lack of risk adjustment.
    - MAP also discussed the current limitations associated with implementing eCQMs and suggests that feasibility testing demonstrates the data are readily available and can be captured without undue burden.
    - MAP also noted there may be a need for balancing measures for cesarean rates (for appropriate populations).
    - MAP suggests that multiple stakeholders including methodological, clinical, and policy experts examine risk adjustment, exclusions, and potential unintended consequences of measuring and reporting cesarean birth rates. Finally, MAP suggested this measure be removed from the HAI domain.

***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 was not clear if those being measured have received feedback on their performance
- Data should be publicly reported.
- Measure is publicly reported

4a. Usability

- As noted, a balancing measure may be needed to ensure the measure does not drive unintended consequences
- Benefits very clear and outweigh any potential harms.
- Transparency of performance and understanding of how care could have looked differently is beneficial to changing practice patterns.

Criterion 5: Related and Competing Measures

Related measures

- NQF #0471 *PC-02 Cesarean Birth*

Harmonization

- The developer is responsible for the maintenance of both the “paper” and “e” versions of the cesarean birth measures.
- The developer attests that NQF #0471e is harmonized with its “paper” version, #0471. The eCQM version of PC02 was developed to reduce administrative burden for sites able to report it and to

encourage the use of eCQMs. The Joint Commission accreditation program accepts either eCQM or chart-abstracted data (or both). Thirteen Joint Commission accredited hospitals submitted PC-02 data for both the eCQM and chart-abstracted measures in calendar year 2020. The ePC-02 rates for the 13 hospitals who submitted both eCQM and chart-abstracted measure results to The Joint Commission for 2020 discharges were correlated at 0.88 which is strong and is statistically significant ($p<0.01$). The eCQM data for this correlation came from 2 EHR systems EPIC and Meditech. The eCQM can decrease burden of manual abstraction, increase efficiencies, and improve experience.

Committee Pre-evaluation Comments:

5: Related and Competing Measures

- PC-02; developer found results to be correlated at 0.88 (strong relationship & stat significant)
- This measure should eventually replace the chart abstraction/paper approach used in NQF #0471 PC-02.
- No

Public and NQF Member Comments (Submitted as of June 10, 2022)

Member Expression of Support

- No public comments received.

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 0471e

Measure Title: ePC-02 Cesarean Birth

Measure is:

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

RELIABILITY: SPECIFICATIONS

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

   Submission document: Items sp.01-sp.30

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

   Reviewer 2: None

   Reviewer 3: None

   Reviewer 5: None

   Reviewer 7: No concerns

   Reviewer 8: There are a number of data sources cited without a clear sense of the data integration required to identify target numerator and denominator.

   Reviewer 9: No concerns

   Reviewer 11: None

RELIABILITY: TESTING

Type of measure:

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

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

Data Source:

☐ **Claims** ☒ **eCQM (HQMF) implemented in EHRs** ☒ **Abstracted from Electronic Health Records**
☐ **Abstracted from Paper Medical Records** ☐ **Instrument-Based Data** ☐ **Registry**
☐ **Enrollment Data** ☐ **Other (please specify)**

Level of Analysis:

☒ **Group/Practice** ☐ **Individual Clinician** ☒ **Hospital/facility/agency** ☐ **Health Plan**
☐ **Population: Regional, State, Community, County or City** ☐ **Accountable Care Organization**
☐ **Integrated Delivery System** ☐ **Other (please specify)**

Submission document: Questions 2a.01-09

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 3:** No hospital level reliability data provided. Patient level data from EHR used to construct measure was assessed for reliability based on comparison with gold standard chart review. Some analysis across hospitals was provided (see Figure 2b.06.01 ePC02 Funnel Plot: 2020 Data, on page 28-29, assessing ability to detect outliers; and Figure 2b.21.02 Observed and Expected rate of 153 California non-best performing hospitals had their patients delivered in the 54 best performing facilities, which constructs an observed to expected rate for 153 CA hospitals show wide variance in performance) but no analysis of consistency of measure over time to assess signal to noise or comparable measure).
   - **Reviewer 6:** Kappa, sensitivity, and specificity analysis was used to assess data validity (and hence data reliability).
   - **Reviewer 7:** Used descriptive statistics to describe hospitals. Data element reliability was done using re-abstraction of data elements compared to EHR extracted/present data. No facility level reliability testing completed as far as I can tell.
   - **Reviewer 9:** It would have been helpful if the tables in this submission were labeled correctly and could stand alone so fewer assumptions would be needed.
   - **Reviewer 10:** The form references the data element validity testing (EHR vs. chart abstracted results). There was no assessment of entity level reliability testing. Not all data elements were assessed, rendering this approach incomplete.
   - **Reviewer 11:** Although sponsor reports validity testing for data elements, it appears as though this was really reliability testing. Chart extraction of data elements to document agreement between what was reported and what was in the chart.

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

   Submission document: Question 2a.11

   - **Reviewer 3**: Data elements drawn from Epic EHR closely match chart review data. There was substantial difficulty in obtaining data from a different EHR system. Ability to obtain accurate data across all major EHR systems not demonstrated. Hospital level analysis does not demonstrate consistency of variations in performance across hospitals.
   - **Reviewer 6**: Sensitivity of measure numerator was a function of testing site. One pilot site had a 88% sensitivity while the 2nd had a 0% sensitivity. Thus, differences in performance across facilities could result from differences in coding accuracy, and not differences in true performance.
   - **Reviewer 7**: Test sample on the smaller side. Use of structured data fields (when able).
   - **Reviewer 9**: Reliability and validity measures were combined. So I'm unsure how to score.
   - **Reviewer 10**: See above, results not included for all data elements, therefore incomplete.
   - **Reviewer 11**: Seven hospitals were tests--6 used Epic and 1 used Meditech but did not enter data regarding birth history or expected date of delivery into the electronic record. Therefore, even though overall there was a high accuracy, this was driven by the 6 Epic hospitals and the available data from the Meditech hospital (which had an overall kappa of 0.477. Similarly 6/32 exclusions were not accurate in Pilot Site 2.

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 2**: Reliability analysis was not required or conducted.
    - **Reviewer 3**: Our concept of reliability has included not only measurement that can be consistently and reliably constructed, but also that performance measured at one point in time or in one sample of patients is likely to be similar to that for a second, close period of time or in another sample of patients. In data presented here, there is no demonstration (1) that the relevant data can be obtained from multiple EHR systems, or (2) that hospital level rates are consistent over time or samples.

- **Reviewer 4:** Results demonstrated overall good accuracy for the identification of numerator events at pilot site 1 (6 hospitals) but poor accuracy at the stand-alone hospital (pilot site 2). Among 6 numerator events detected by manual review at pilot site #2, none of them were captured as numerator events by EHR. This illustrates the potential for errors in EHR-based data capture and suggests that accuracy results may be heavily site-specific. If issues can be tested and resolved for each site that participates, then it's reasonable to assume the elements will be captured accurately.
- **Reviewer 5**: See validity
- **Reviewer 6**: Sensitivity of measure numerator was a function of testing site. One pilot site had a 88% sensitivity while the 2nd had a 0% sensitivity. Thus, differences in performance across facilities could results from differences in coding accuracy, and not differences in true performance.
- **Reviewer 7**: score level testing?
- **Reviewer 8:** There was substantial between hospital variation, particularly across sites, however the sample size is very small, denominators vary substantially and within hospital variance was not estimated, there were a number of outliers (without risk adjustment) and at least one hospital outside the US. Data presented to not provide compelling evidence of accountable-entity reliability.
- **Reviewer 9**: Because reliability and validity were presented together I am unsure how to score.
- **Reviewer 10**: Insufficient data provided for data element validity No data for measure exclusions (Inpatient hospitalizations for patients with abnormal presentation or placenta previa during the encounter)
- **Reviewer 11**: Testing highlights fact that many hospitals may have systems that are not adequate for reliance on electronic transmission of data for this measure.

VALIDITY: TESTING

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

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

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

    Submission document*:* Questions 2b.01-02.

    ☒ **Yes**

    ☒ **No**

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

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

    NOTE: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

    Submission document: Questions 2b.01-02

    ☐ **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**: Review of charts and comparison to electronic medical record to calculate agreement rates (kappa statistics). Review of performance scores at hospital level showed wide variation in performance.
- **Reviewer 2**: Item-level validity was conducted by comparing e-extracted data elements to chart review gold standard
- **Reviewer 3**: Data element validity from EHRs tested by comparing to chart review.
- **Reviewer 5**: Sample sizes are very small
- **Reviewer 6**: Kappa, sensitivity, and specificity analysis.
- **Reviewer 7**: Abstraction. PPV.
- **Reviewer 9**: Method was chart review. Seems acceptable.
- **Reviewer 10**: The developer used kappa to assess agreement across 101 patients in 7 hospitals. Kappa reported for the outcome, not each data element in the measure.
- **Reviewer 11**: There was a correlation performed relating to anther measure (birth complications). However this analysis was used to variation in this measure compared to a predicted score based on the performance of top performing hospitals adjusted for BMI and age. However, no direct face or empirical testing of measure validity was presented.

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

Submission document: Questions 2b.03-04

- **Reviewer 1**: The overall kappa for data element agreement was 0.955 which indicates excellent agreement. The exceptions to this were the secondary diagnoses (other than Single Live Term Newborn) and the procedure codes which were lower since they were not always collected according to the instructions. In addition, the gestational age, author datetime, and birth weight result agreement rates were low due to differing data sources. The demographic variables of race and ethnicity also had lower agreement rates for site 2 which were due to different data sources. Despite the lower agreement rates for these noted data elements, measure scores validity was not impacted per the developer.

- Measure score validity was high as measured by the kappa, sensitivity, specificity, PPV and NPV statistics. All of these statistics were greater than 0.90. No further information was provided except this brief summary.
- **Reviewer 2**: The item-level validity analysis (e-extraction to chart review gold standard) was good for Site 1 and not for Site 2. The average performance metrics are not useful because the measure still isn't working very well in Site 2 which uses Meditech. Its unclear how many, if any, of the Site 1 hospital use Meditech or if they are all EPIC. Is it concerning if the measure seems valid only for one EHR. Is there an expectation that measures are tested in multiple EHRs and are valid in each?
- **Reviewer 3**: Data from hospitals with EPIC EHRs appeared sufficient. Data from second EHR was not reliable, raising validity issues.
- **Reviewer 5**: Very few sites tested
- **Reviewer 6**: Sensitivity of measure numerator was a function of testing site. One pilot site had a 88% sensitivity while the 2nd had a 0% sensitivity. Thus, differences in performance across facilities could results from differences in coding accuracy, and not differences in true performance.
- **Reviewer 7**: Adequate for one EHR (EPIC). Should test on other EHRs for validation of those hospitals that don't use EPIC.

- **Reviewer 8:** Although the new format of the submission makes it difficult to tell, it appears that kappa for data element reliability (highly variable by site) was also used to assess validity. PPV for HER vs. manual chart abstraction is again data element validity vs. accountable-entity validity. Funnel plot results are not compelling and violate many of the basic assumptions of this strategy for assessing quality of care measures (see below): We developed these guidelines following a focused literature search, in which we identified six conceptual steps in constructing a funnel plot. These are: (1) defining policy level input; (2) checking the quality of models used for case-mix correction; (3) examining whether the number of observations per hospital is sufficient to fulfill the assumptions upon which the control limits are based; (4) testing for overdispersion of the values of the quality indicator; (5) testing whether the values of the quality indicators are associated with institutional characteristics; and (6) specifying how the funnel plot should be constructed. A funnel plot can be used as a tool to identify a small percentage of deviating institutions. It is not meant to beused to judge whether different groups of institutions perform differently. For this reason, quality indicators can only be validly presented in funnel plots if there is no association between the values of the quality indicator and hospital characteristics.3 Ideally, differences between hospitals only represent true differences in the quality of care and randomvariations.33 However, overdispersion occurs when there is true heterogeneity between hospitals, over andabove that expected due to random variation.33–41 If overdispersion occurs, one needs to be careful to draw conclusions from the funnel plot, since the assumptions with respect to the distribution of the qualityindicator are violated. Often the cause of overdispersion is not clear, but heterogeneity may arise when hospitals serve patients with different characteristics for which the model does not sufficiently correct; due to registration bias or errors; or policy choices or variability in the actual quality of care offered.39 or policy choices or variability in the actual quality of care offered.39 The Brier score is a mixture between discrimination and calibration and can range from 0 for a perfect model to 0.25 for a non-informative model with a 50% incidence of the outcome. The Brier score could be scaled by its maximum score, which is lower if the incidence is lower. https://journals.sagepub.com/doi/pdf/10.1177/0962280217700169
- Tests for over dispersion were not provided, nor were Brier scores
- **Reviewer 9**: Site 2 had some real problems. If this happens in practice - how will the site know they have data issues? will they have misclassification due to these issues?
- **Reviewer 10**: Numerator Kappa agreement at one site was good at .831, while the 2nd site was poor at .477. however, not all data elements were tested for chance-adjusted agreement. At site 2, the developer notes: No numerator events were identified based on submitted data. Six numerator events were evaluated during validity testing. During validity testing six cases that were originally excluded from the denominator due to a missing time of delivery were found to meet the numerator when time of delivery was provided. This hospital uses a standalone OB documentation system that does not interface completely with the electronic health record (Meditech). The OB documentation is present in Meditech in non-discrete fields in a .pdf form. The kappa results were not provided for the denominator exclusions.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

    Submission document: Questions 2b.15-18.

- **Reviewer 1**: They said because number of sites was small, no formal statistical test was performed for the effect of exclusion on the performance scores. This is concerning. It is not clear where there are large numbers of exclusions whether patients with social risk factors were affected more by exclusions and whether this would inappropriately inflate the facility performance?
- **Reviewer 2:** None
- **Reviewer 3:** None

- **Reviewer 5:** None
- **Reviewer 7:** None
- **Reviewer 9:** No concerns
- **Reviewer 10:** None

19. **Risk Adjustment**

Submission Document: Questions 2b.19-32

19a. **Risk-adjustment method**

☒ None ☐ 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 1**: The developers claim the exclusions result in a homogenous population that represents the group where most improvement opportunity exists. The variation in rates suggest differences due to clinical practice patterns and thus room for improvement. Clinicians will need to comment on whether there are other appropriate exclusions or comorbidities/characteristics that should be risk adjusted.
- **Reviewer 2**: I view this as a process measures and agree that it should not be risk adjusted. The exclusions were used to make sure that everyone in the denominator should equally avoid Cesarean delivery.
- **Reviewer 3**: Some empirical analysis of additional exclusion rules and age/BMI analysis supports decision not to risk adjust.
- **Reviewer 5**: Yes
- **Reviewer 6**: The measure is not risk adjusted. The MD should consider the use of maternal age, BMI, pre-existing comorbid conditions, obstetrical conditions (pre-eclampsia), and prolonged labor. The MD provide evidence that university-based hospitals (who presumably have higher risk patients) have similar CD rates compared to other sites. This, however, does not provide sufficient evidence to demonstrate that some of the variation in CD rates would differ in individual hospitals with very degrees of patient risk.
- **Reviewer 7**: No analysis done on minority women other than to say small sample size but sufficiently varied. No data to back up claim.

- **Reviewer 8:** There is not compelling rationale for failing to risk adjust this measure, as recommended by the funnel plot guidelines associated with the references noted about.
- **Reviewer 9**: No concerns.
- **Reviewer 11**: The only risk factors explored are age and BMI.

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

    Submission document: Questions 2b.05-07

    - **Reviewer 3**: Provided data shows ability to identify variations in performance.
    - **Reviewer 7**: As noted above.
    - **Reviewer 9**: 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.

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

    Submission document: Questions 2b.08-10.

    - **Reviewer 1**: One of the pilot sites had extensive data issues/missing data. This seemed to be related to the electronic health record system utilized which is concerning, but this site has a plan to remediate.
    - **Reviewer 3:** Non EPIC EHR did not always contain relevant data
    - **Reviewer 7:** No information on how missing data is handled. Seems like its just not sought after or included. Missing data specific to certain sites and certain data elements.
    - **Reviewer 8:** The overall missing data rate based on Table 2b.09.01 appears to have missing data rates >10% in many important variables and variable by Pilot Site, raising concerns about the impact of missing data on results.
    - **Reviewer 9:** Missing data seems to be site specific. How can this be addressed?
    - **Reviewer 11:** There was identified a systemic source of missing data in one of the 7 hospital used for data validity testing. Even though sponsored mentioned possible amelioration plan for this hospital it is not clear that this problem would not occur in other hospitals

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**: No testing of exclusions due to sample size. Measure may not be ready for implementation without further testing.
    - **Reviewer 2:** The item-level validity analysis (e-extraction to chart review gold standard) was good for Site 1 and not for Site 2. The average performance metrics are not useful because the measure still isn't working very well in Site 2 which uses Meditech. Its unclear how many, if any, of the Site 1 hospital use Meditech or if they are all EPIC. Is it concerning if the measure seems valid only for one EHR. Is there an expectation that measures are tested in multiple EHRs and are valid in each? In any event, I am not confident that the validity results for subsequent sites will mirror Site 1.
    - **Reviewer 3:** Need to know that all relevant data elements can in fact be obtained from EHR.
    - **Reviewer 4:** Developers argue that risk-adjustment is not required because the population is relatively homogeneous risk and factors that make patients high risk for needing cesarean birth do not vary systematically across hospitals. The exclusion criteria were chosen intentionally to yield a relatively homogeneous population which would not require further risk adjustment. They argue that the impact of risk factors is small relative to the magnitude of true quality differences.
    - **Reviewer 7:** Data element validity via manual abstract comparison with EHR data collected.
    - **Reviewer 8:** The facility level validity evidence relies primarily on data element "validity" which is consistency/agreed across data sources (reliability/consistency). Facility level validation with exogenous variables does not appear to have been performed. The small number of facilities and pilot site/facility variation further limit conclusions on measure validity based on the data provided.
    - **Reviewer 10:** Lacks kappa agreement for exclusions.
    - **Reviewer 11:** In the absence of documentation of face validity or of agreement with other measures generally accepted as related to quality the validity of the measure remains to be determined.

For composite measures ONLY

Submission documents: Questions 2c.01-08

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**

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.**
    - **Reviewer 6:** 1. Lack of risk adjustment. 2. Data element validity was only demonstrated in one of the 2 test sites.

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

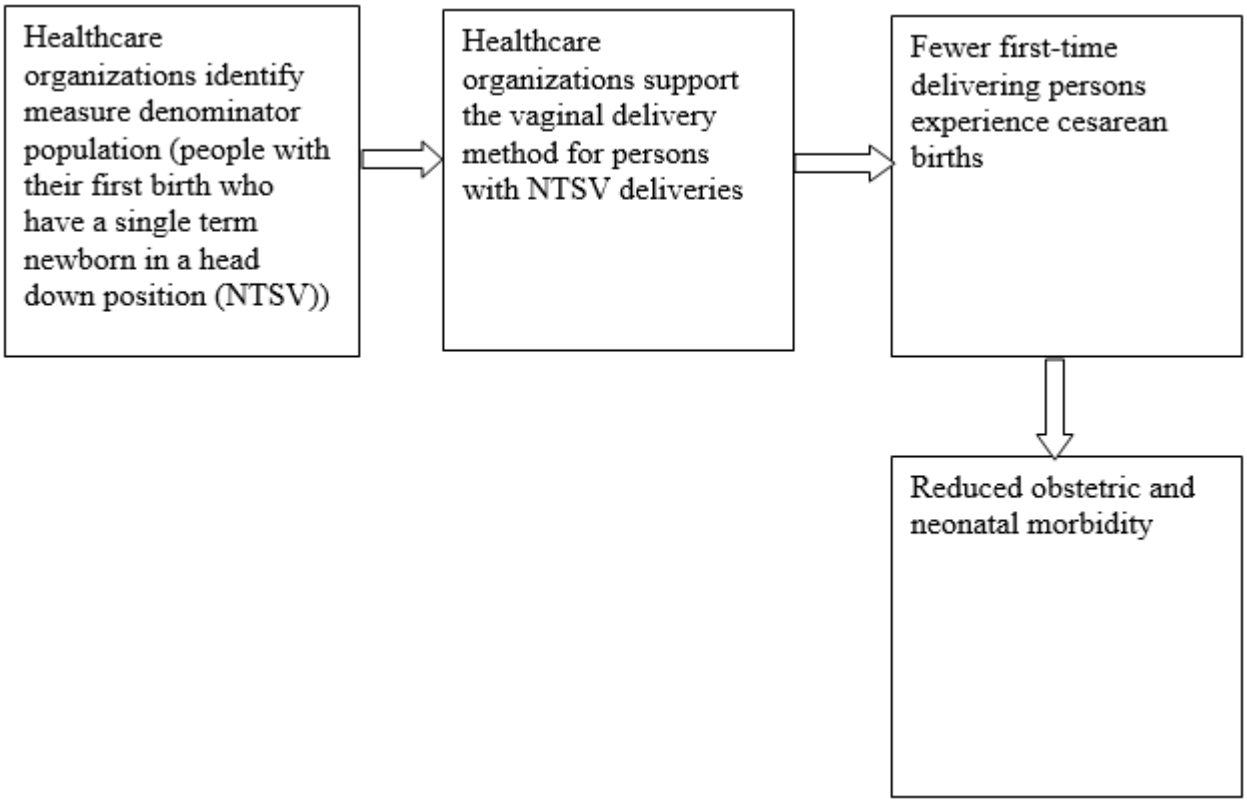
---

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 measure will assist health care organizations to track nulliparous patients with live term singleton newborns in vertex position delivering by cesarean birth to reduce the occurrence of cesarean births.
A reduction in the number of nulliparous patients with live term singleton newborns in vertex position (NTSV) delivering by cesarean birth will result in increased patient safety, a substantial decrease in maternal and neonatal morbidity and substantial savings in health care costs. Successful quality improvement efforts incorporate audit and feedback strategies combined with provider and nurse education, guidelines and peer review.

[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]
Sakala et al., (2020) performed a secondary analysis of the *Listening to Mothers in California Survey.* The study provided outreach materials and a survey questionnaire which were to address California population and policy issues. A sample of women who had a cesarean delivery between September 1 and December 15, 2016, were selected for participation. Relevant survey results showed almost one in three (30%) respondents sought information about cesarean rates of prospective hospitals, with the majority able to find this information. Also, about one in three correctly understood that the quality of care varies across both obstetricians and hospital maternity units. Providing women with a reliable resource to find hospital cesarean birth rates, supports the engagement of women in their care. This study provides evidence that the target population values the measured outcome, and the quality of the care they receive.

- Sakala, C., Belanoff, C., & Declercq, E. R. (2020). Factors Associated with Unplanned Primary Cesarean Birth: Secondary Analysis of the Listening to Mothers in California Survey. BMC pregnancy and childbirth, 20(1), 462. https://doi.org/10.1186/s12884-020-03095-4

[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]
American College of Obstetricians and Gynecologists (College), Society for Maternal-Fetal Medicine, Caughey, A. B., Cahill, A. G., Guise, J. M., & Rouse, D. J. (2014, reaffirmed 2019). Safe prevention of the primary cesarean delivery. *American journal of obstetrics and gynecology, 210*(3), 179–193. https://doi.org/10.1016/j.ajog.2014.01.026
The intent of this measure is to reduce the rates of cesarean sections, and as a result, reduce the number of complications and negative outcomes for moms and babies. In 2011, one in three women who gave birth in the United States did so by cesarean delivery. Even though the rates of primary and total cesarean delivery have plateaued recently, there was a rapid increase in cesarean rates from 1996 to 2011. Although cesarean delivery can be lifesaving for the fetus, the mother, or both in certain cases, the rapid increase in the rate of cesarean births without evidence of concomitant decreases in maternal or neonatal morbidity or mortality raises significant concern that cesarean delivery is overused. Therefore, it is important for health care providers to understand the short-term and long-term tradeoffs between cesarean and vaginal delivery, as well as the safe and appropriate opportunities to prevent overuse of cesarean delivery, particularly primary cesarean delivery.
The recommendations by the American College of Obstetricians and Gynecologists have identified 40 studies that help to support the evidence in reducing the number of nulliparous, term, singleton newborns in vertex position to be delivered by cesarean birth. Of those studies, there were Population-based, Retrospective-Cohort, Prospective, Randomized trials, and Meta-analysis studies performed. Some of those recommendations include: allowing for increased length of time for pushing during the active stage of labor (at least 2 hours of pushing in multiparous women and 3 hours of pushing for nulliparous patients) and operative vaginal delivery in the second stage of labor by experienced and well-trained physicians should be considered a safe, acceptable alternative to cesarean delivery. Training in, and ongoing maintenance of, practical skills related to operative vaginal delivery should be encouraged (forceps or manual rotation/aversion). Several approaches are needed to reduce the primary cesarean delivery rate, which in turn would lower the repeat cesarean delivery rate. Although national and regional organizations can take the lead in setting the agenda regarding the safe prevention of primary cesarean delivery, such an agenda will need to be prioritized at the level of practices, hospitals, health care systems, and, of course, patients.
A 2007 review found that the cesarean delivery rate was reduced by 13% when audit and feedback were used exclusively and decreased by 27% when audit and feedback were used as part of a multifaceted intervention, which involved second opinions and culture change. Systemic interventions, therefore, provide an important strategic opportunity for reducing cesarean delivery rates. However, the specific interventional approaches have not been studied in large, prospective trials, thus specific recommendations cannot be made.

As noted in this source, a large population-based study from Canada found that the risk of severe maternal morbidities—defined as hemorrhage that requires hysterectomy or transfusion, uterine rupture, anesthetic complications, shock, cardiac arrest, acute renal failure, assisted ventilation, venous thromboembolism, major infection, or in-hospital wound disruption or hematoma—was increased threefold for cesarean delivery as compared with vaginal delivery (2.7% versus 0.9%, respectively). There also are concerns regarding the long-term risks associated with cesarean delivery, particularly those associated with subsequent pregnancies. The incidence of placental abnormalities, such as placenta previa, in future pregnancies increases with each subsequent cesarean delivery, from 1% with one prior cesarean delivery to almost 3% with three or with three or more prior cesarean deliveries. This combination of complications not only significantly increases maternal morbidity but also increases the risk of adverse neonatal outcomes, such as neonatal intensive care unit admission and perinatal death.

A cross-sectional study of the 2015–2017 California Maternal Quality Care Collaborative (CMQCC) statewide collaborative was conducted to support vaginal birth and reduce primary cesarean delivery. The study solicited hospitals with nulliparous, term, singleton, vertex cesarean delivery rates greater than 23.9%. Fifty-six hospitals with more than 119,000 annual births participated. Of these, 87.5% were community facilities. Safety measures were derived using data collected as part of routine care and submitted monthly. Data was obtained from birth certificates, maternal and neonatal discharge diagnosis and procedure files, and selected clinical data elements submitted as supplemental data files. Maternal measures included chorioamnionitis, blood transfusions, third- or fourth-degree lacerations, and operative vaginal delivery. Neonatal measures included the severe unexpected newborn complications metric and 5-minute Apgar scores less than 5. Mixed-effect multivariable logistic regression model was used to calculate odds ratios (Ors) and 95% CIs.

Results demonstrated among collaborative hospitals that the nulliparous, term, singleton, vertex cesarean delivery rate fell from 29.3% in 2015 to 25.0% in 2017. The tercile of hospitals with the greatest decline (31.2%–20.6%, 2017 vs 2015 aOR 0.54, 95% CI 0.50–0.58) was evaluated to determine whether they had greater risk of poor maternal and neonatal outcomes. No measure was statistically worse, and the severe unexpected newborn complications composite actually declined (3.2%–2.2%, aOR 0.71, 95% CI 0.55–0.92). The findings of this study support that reductions in cesarean delivery rates need not lead to worse neonatal or maternal outcomes.

Main, E. K., Chang, S. C., Cape, V., Sakowski, C., Smith, H., & Vasher, J. (2019). Safety Assessment of a Large-Scale Improvement Collaborative to Reduce Nulliparous Cesarean Delivery Rates. *Obstetrics and gynecology*, *133*(4), 613–623. https://doi.org/10.1097/AOG.0000000000003109

In 2015, nulliparous, term, singleton, vertex cesarean delivery rate ranged from 11.3% to 76.9% in 248 California hospitals with maternity services. Fifty-six hospitals, all among those with initial rates above 23.9% participated in California Maternal Quality Care Collaborative's (CMQCC) Supporting Vaginal Birth collaborative. The overall nulliparous, term, singleton, vertex cesarean delivery rates declined by 4.5 percentage points (a relative decline of 15.5%), from 29.1% in the first two quarters in 2015 to 24.6% in the last two quarters in 2017. The largest decline happened in the third and fourth quarters of 2016, when the collaborative initiated. Twenty-four of the 56 participating hospitals lowered their nulliparous, term, singleton, vertex cesarean delivery rates to below the Healthy People 2020 target of 23.9% in 2017. Ten of them lowered the rates even further, to below 20% (range 15.0–19.9%). Similar to the analysis based on the absolute amount of decline shown above, we did not observe an increase in adverse maternal or neonatal outcomes even among those with these lower final rates (hospitals with 2017 nulliparous, term, singleton, vertex cesarean delivery rate between 15.0% and 19.9%). Rates of severe unexpected newborn complications also declined from 2.5% in 2015 to 2.2% in 2017 in this group, but not significantly with an adjusted OR of 0.84 (95% CI 0.58–1.20). The size of this study is noteworthy. Fifty-six hospitals participated with a total annual delivery volume of 119,000 women. This is a higher delivery volume than in all but nine U.S. states. All hospitals in the collaborative had a starting nulliparous, term, singleton, vertex cesarean delivery rate higher than the Healthy People 2020 national target of 23.9%.18 Importantly, the majority of collaborative hospitals were community hospitals (87%), representing the predominant care model in the United States. The CMQCC Supporting Vaginal Birth collaborative interventions emphasized reducing latent phase cesarean deliveries, implementation of ACOG–SMFM guidelines for diagnosis and management of active phase disorders, and enhanced nursing support (increased walking and upright positioning, use of peanut balls, and interpersonal coaching). Providers did not increase their operative vaginal delivery rates. Findings provide evidence that a reduction in first birth cesarean delivery rates need not be associated with more difficult vaginal births or higher rates of major perineal lacerations. The large number of nulliparous, term, singleton, vertex deliveries provided the ability to confidently examine maternal and neonatal complications that are infrequent. The range of improvement among the 56 hospitals created the opportunity to perform a sensitivity analysis comparing hospitals with very high levels of cesarean delivery rate reduction (217.1 to 27.1 percentage points) with those with limited change (22.4 to +4.7 percentage points). The rate of severe unexpected newborn complications (the major composite index for neonatal outcomes) actually improved in hospitals with the greatest reduction in nulliparous, term, singleton, vertex cesarean delivery rate. This is in concordance with several of the single hospital studies noted.

[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 removal of any pressure to not perform a cesarean birth has led to a skyrocketing of hospital, state and national cesarean birth (CB) rates. Some hospitals' CB rates were over 50%. Hospitals with CB rates at 15-20% have infant outcomes that are just as good and better maternal outcomes (Symum et al., 2021). There is no data that higher rates improve any outcomes, yet the CB rates continue to rise. This measure seeks to focus attention on the most variable portion of the CB epidemic, the term labor CB in nulliparous women. This population segment accounts for the large majority of the variable portion of the CB rate and is the area most affected by subjectivity.
As compared to other CB measures, what is different about NTSV CB rate (Primary CB in first births with term singleton pregnancies in head down position) is that there are clear cut quality improvement activities that can be done to address the differences. Main et al. (2012) found that over 60% of the variation among hospitals can be attributed to first birth labor induction rates and first birth early labor admission rates. The results showed if labor was forced when the cervix was not ready the outcomes were poorer. Rosenstein et al. (2021) also showed that labor and delivery guidelines can make a difference in labor outcomes. Many authors have shown that physician factors, rather than patient characteristics or obstetric diagnoses are the major driver for the difference in rates within a hospital (Berkowitz, et al.,1989; Goyert et al., 1989; Luthy et al., 2003, Symum et al., 2021). The dramatic variation in cesarean rates seen in all populations studied is striking. (Cesarean rates varied tenfold in US hospitals nationwide across hospitals, from 7.1 % to 69.9 % and there was a 15-fold variation among low-risk women, from 2.4% to 36.5% (Kozhimannil et al., 2013).
A reduction in the number of nulliparous patients with live term singleton newborns in vertex position (NTSV) delivering by cesarean birth will result in increased patient safety, a substantial decrease in maternal and neonatal morbidity and substantial savings in health care costs. Successful quality improvement efforts incorporate audit and feedback strategies combined with provider and nurse education, guidelines and peer review.
The measure will assist health care organizations (HCOs) to track nulliparous patients with live term singleton newborns in vertex position delivering by cesarean birth to reduce the occurrence. Nulliparous women have 4-6 times the cesarean birth rate than multiparous women thus the NTSV population is the largest driver of primary cesarean birth rate (Sakala et al. 2020). NTSV has a large variation among facilities, thus identifying an important population on which to focus quality improvement efforts.
In addition, a reduction in primary cesarean births will reduce the number of women having repeat cesarean births (almost 90% of mothers who have a primary cesarean birth will have subsequent cesarean birth (CDC, 2020)). Thus, improvement in the rates of cesarean birth for the first birth will reduce the morbidity of all future births and avoid all the controversies with trial of labor after cesarean/elective repeat cesareans.
Sources

- Agency for Healthcare Research and Quality. (2002). AHRQ Quality Indicators Guide to Inpatient Quality Indicators: Quality of Care in Hospitals Volume, Mortality, and Utilization. Revision 4 (December 22, 2004). AHRQ Pub. No. 02-RO204.
- American College of Obstetricians and Gynecologists. (2000). Task Force on Cesarean Delivery Rates. Evaluation of Cesarean Delivery. (Developed under the direction of the Task Force on Cesarean Delivery Rates, Roger K. Freeman, MD, Chair, Arnold W. Cohen, MD, Richard Depp III, MD, Fredric D. Frigoletto Jr, MD, Gary D.V. Hankins, MD, Ellice Lieberman, MD, DrPH, M. Kathryn Menard, MD, David A. Nagey, MD, Carol W. Saffold, MD, Lisa Sams, RNC, MSN and ACOG Staff: Stanley Zinberg, MD, MS, Debra A. Hawks, MPH, and Elizabeth Steele)
- Bailit, J.L., Garrett, J.M., Miller, W.C., McMahon, M.J., & Cefalo, R.C. (2002). Hospital primary cesarean delivery rates and the risk of poor neonatal outcomes. Am J Obstet Gynecol. 187(3):721-7.
- Bailit, J. & Garrett, J. (2003). Comparison of risk-adjustment methodologies. Am J Obstet Gynecol.102:45-51.
- Bailit, J.L., Love, T.E., & Dawson, N.V. (2006). Quality of obstetric care and risk-adjusted primary cesarean delivery rates. Am J Obstet Gynecol.194:402.
- Bailit, J.L. (2007). Measuring the quality of inpatient obstetrical care. Ob Gyn Sur. 62:207-213.
- Berkowitz, G.S., Fiarman, G.S., Mojica, M.A., et al. (1989). Effect of physician characteristics on the cesarean birth rate. Am J Obstet Gynecol. 161:146-9.

- Cleary, R., Beard, R.W., Chapple, J., Coles, J., Griffin, M., & Joffe, M. (1996). The standard primipara as a basis for inter-unit comparisons of maternity care. Br J Obstet Gynecol. 103:223-9.
- DiGiuseppe, D.L., Aron, D.C., Payne, S.M., Snow, R.J., Dieker, L., & Rosenthal, G.E. (2001). Risk adjusting cesarean delivery rates: a comparison of hospital profiles based on medical record and birth certificate data. Health Serv Res.36:959-77.
- Goyert, G.L., Bottoms, F.S., Treadwell, M.C., et al. (1989). The physician factor in cesarean birth rates. N Engl J Med.320:706-9.
- Kozhimannil, K. B., Law, M. R., & Virnig, B. A. (2013). Cesarean delivery rates vary tenfold among US hospitals; reducing variation may address quality and cost issues. Health affairs (Project Hope), 32(3), 527–535. https://doi.org/10.1377/hlthaff.2012.1030
- Le Ray, C., Carayol, M., Zeitlin, J., Berat, G., & Goffinet, F. (2006). Level of perinatal care of the maternity unit and rate of cesarean in low-risk nulliparas. Am J Obstet Gynecol. 107:1269-77.
- Luthy, D.A., Malmgren, J.A., Zingheim, R.W., & Leininger, C.J. (2003). Physician contribution to a cesarean delivery risk model. Am J Obstet Gynecol.188:1579-85.
- Main E.K., Bloomfield, L., & Hunt, G. (2004). Development of a large-scale obstetric quality-improvement program that focused on the nulliparous patient at term. Am J Obstet Gynecol.190:1747-58.
- Main, E. K., Chang, S. C., Cape, V., Sakowski, C., Smith, H., & Vasher, J. (2019). Safety Assessment of a Large-Scale Improvement Collaborative to Reduce Nulliparous Cesarean Delivery Rates. Obstetrics and gynecology, 133(4), 613–623. https://doi.org/10.1097/AOG.0000000000003109
- Main, E.K., Moore, D., Farrell, B., Schimmel, L.D., Altman, R.J., Abrahams, C., et al., (2006). Is there a useful cesarean birth measure? Assessment of the nulliparous term singleton vertex cesarean birth rate as a tool for obstetric quality improvement. Am J Obstet Gynecol. 194:1644-51.
- Main, E. K., Morton, C. H., Melsop, K., Hopkins, D., Giuliani, G., & Gould, J. B. (2012). Creating a public agenda for maternity safety and quality in cesarean delivery. Obstetrics and gynecology, 120(5), 1194–1198. https://doi.org/10.1097/aog.0b013e31826fc13d
- Center for Disease Control (2020) *Recent trends in vaginal birth after cesarean delivery: United States, 2016-2018*. Retrieved from National Center for Health Statistics: https://www.cdc.gov/nchs/products/databriefs/db359.htm
- Romano, P.S., Yasmeen, S., Schembri, M.E., Keyzer, J.M., & Gilbert, W.M. (2005). Coding of perineal lacerations and other complications of obstetric care in hospital discharge data. Am J Obstet Gynecol.106:717-25.
- Rosenstein, M. G., Chang, S. C., Sakowski, C., Markow, C., Teleki, S., Lang, L., Logan, J., Cape, V., & Main, E. K. (2021). Hospital Quality Improvement Interventions, Statewide Policy Initiatives, and Rates of Cesarean Delivery for Nulliparous, Term, Singleton, Vertex Births in California. *JAMA*, *325*(16), 1631–1639. https://doi.org/10.1001/jama.2021.3816
- Sakala, C., Belanoff, C., & Declercq, E. R. (2020). Factors Associated with Unplanned Primary Cesarean Birth: Secondary Analysis of the Listening to Mothers in California Survey. BMC pregnancy and childbirth, 20(1), 462. https://doi.org/10.1186/s12884-020-03095-4
- Symum, H., & Zayas-Castro, J. L. (2021). A Multistate Decomposition Analysis of Cesarean Rate Variations, Associated Health Outcomes, and Financial Implications in the United States. *American journal of perinatology*, 10.1055/s-0041-1736538. Advance online publication. https://doi.org/10.1055/s-0041-1736538
- U.S. Department of Health and Human Services. (n.d.). Reduce cesarean births among low-risk women with no prior births-MICH-06. Retrieved from Healthy People 2030: https://health.gov/healthypeople/objectives-and-data/browse-objectives/pregnancy-and-childbirth/reduce-cesarean-births-among-low-risk-women-no-prior-births-mich-06
- Yasmeen, S., Romano, P.S., Schembri, M.E., Keyzer, J.M., & Gilbert, W.M. (2006). Accuracy of obstetric diagnoses and procedures in hospital discharge data. Am J Obstet Gynecol. 194:992-1001.

[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]

For the pilot study, there were a limited number of hospitals, therefore the five number statistical summaries are used in place of deciles. Improvement is noted as a decreasing trend. The Joint Commission does not want to encourage inappropriately low Cesarean rates that may be unsafe to patients. Acceptable PC-02 rates are 30% or lower, however there is not an established threshold for what rate may be too low. A rate greater than 30% is about two standard deviations from the mean two-year rate and The Joint Commission has set this as the threshold used for public reporting of the measure. PC-06 serves as a balancing measure for PC-02 to guard against any unanticipated or unintended consequences and to identify unforeseen complications that might arise as a result of quality improvement activities and efforts for this measure.

Data are summarized at the hospital level for 2020 discharges

Number of hospitals: 15

Data Source: EHR

Median number of denominator cases: 108

Table 1b.02.01 Cesarean Birth Measure Rates

Statistic	Value
Mean	27.5%, SD 20.0%
Min	0%
25th Percentile	19.5%
50th Percentile	23.3%
75th Percentile	28.9%
Max	71.8%

Table 1b.02.01 Cesarean Birth Measure Rates provides the mean (27.5%), minimum (0%) and maximum (71.8%) and the 25th, 50th and 75th percentile for 15 hospitals that submitted data for 2020 discharges.

Table 1b.02.02 Hospital Characteristics

Hospital Characteristic	Values
State Represented	6 states represented and Puerto Rico CA, CT, IA, MI, OH, PR, WA
Hospital Location	12 Urban 3 Rural
Control / Ownership Type	9 Nongovernment (not-for-profit) 3 Government 3 For Profit
Primary Service	14 General medical and surgical 1 Specialty
Teaching Affiliation	11 Minor teaching 4 Non-teaching
System member	10 Yes 5 No
Beds (total facility)	6 <100 beds 5 100-399 beds 4 400 + beds
Total births (per year)	Range 165-5323 6 100-499 4 500-999 4 1000-4999 1 Greater than 5000

Table 1b.02.02 Hospital Characteristics describes the state represented, hospital location, control/ownership type, primary service, teaching affiliation, system member, beds and total births for the 15 hospitals that submitted data for 2020 discharges.

[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]

See 1b.02

The Healthy People 2030 national goal for cesarean births among low-risk women with no prior births is 23.6%. in 2018 the rate was 25.9 and in 2019 it decreased to 25.6. This national rate shows there is still room for improvement. 2020 Individual hospital rates show an average rate of 27.5%.

The ePC-02 measure highly correlates with the chart-abstracted PC-02 measure at 0.88 (see 5.06). Below are the PC-02 chart-abstracted data trends from 2018-2020. Since the eCQM is harmonized with the chart-based version of the measure and highly correlated, we expect similar trends with the eCQM data. There is a decreasing IQR trend from 2018 to 2020, however there is still variation in the performance rates and 20% of hospitals still have a rate above 30% in 2020. A rate greater than 30% is about two standard deviations from the mean two-year rate and The Joint Commission has set this as the threshold used for public reporting of the measure.

1936 hospitals submitted PC-02 chart-abstracted data for CY 2018 which included 490,481 patient records.

Table 1b.03.01 PC-02 Chart-abstracted 2018 Data

PC-02 Chart-abstracted Data	CY 2018 Statistics
Mean (SD)	25.7% (7.6%)
IQR	8.9%
Deciles (0)	2.0%
Deciles (10)	16.9%
Deciles (20)	19.7%
Deciles (30)	21.8%
Deciles (40)	23.5%
Deciles (50)	25.0%
Deciles (60)	26.8%
Deciles (70)	28.6%
Deciles (80)	30.9%
Deciles (90)	34.8%
Deciles (100)	100%

Table 1b.03.01 PC-02 Chart-abstracted 2018 Data shows the mean (25.7%), interquartile range (8.9%), and 0 to 100th deciles. The 50th percentile is 25.0%. 1936 hospitals submitted PC-02 chart-abstracted data for CY 2018 which included 490,481 patient records.

1909 hospitals submitted chart-abstracted PC-02 data for CY2019 which included 491,893 patient records.

Table 1b.03.02 PC-02 Chart-abstracted 2019 Data

PC-02 Chart-abstracted Data	CY 2019 Statistics
Mean (SD)	24.8% (7.2%)
IQR	8.7%
Deciles (0)	4.3%
Deciles (10)	16.7%
Deciles (20)	19.1%
Deciles (30)	21.1%
Deciles (40)	22.70%
Deciles (50)	24.3%
Deciles (60)	26.0%
Deciles (70)	27.90%
Deciles (80)	30.20%
Deciles (90)	33.6%
Deciles (100)	71.4%

Table 1b.03.02 PC-02 Chart-abstracted 2019 Data shows the mean (24.8%), interquartile range (8.7%), and 0 to 100th deciles. The 50th percentile is 24.3%. 1909 hospitals submitted chart-abstracted PC-02 data for CY2019, which included 491,893 patient records.
1214 hospitals submitted PC-02 chart-abstracted data for CY 2020 which included 313,591 patient records.

Table 1b.03.03 PC-02 Chart-abstracted 2020 Data

PC-02 Chart-abstracted Data	CY 2020 Statistics
Mean (SD)	24.9% (6.9%)
IQR	8.5%
Deciles (0)	6.0%
Deciles (10)	16.3%
Deciles (20)	19.0%
Deciles (30)	21.4%
Deciles (40)	23.0%
Deciles (50)	24.8%
Deciles (60)	26.3%
Deciles (70)	27.9%
Deciles (80)	30.2%
Deciles (90)	33.8%
Deciles (100)	68.8%

Table 1b.03.03 PC-02 Chart-abstracted 2020 Data shows the mean (24.9%), interquartile range (8.5%), and 0 to 100th deciles. The 50th percentile is 24.8%. 1214 hospitals submitted PC-02 chart-abstracted data for CY 2020 which included 313,591 patient records.

[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]
The 2020 data provided below represents 993 discharges in the measure population from 15 hospitals. Variability in measure rates demonstrate opportunity for improvement and importance of monitoring for disparities. See Table 1b.02.02 for hospital characteristics.

Table 1b.04.01 Measure Rates by Age Category

Age	rate	N
<20	20.2%	94
20-25	25.0%	276
25-30	30.6%	242
30-35	25.4%	252
35-40	33.7%	101
40+	42.9%	28

Table 1b.04.01 Measure Rates by Age Category displays the number and measure rate for patients in the following age categories: <20 (20.2%), 20-25 (25%), 25-30 (30.6%), 30-35 (25.4%), 35-40 (33.7%), 40 plus (42.9%). The 2020 data represents 993 discharges in the measure population from 15 hospitals.

Table 1b.04.02 Measure Rate by Hispanic Ethnicity

Hispanic ethnicity	rate	N
Hispanic or Latino	35.8%	296
Not Hispanic or Latino	23.9%	695
Missing	0.0%	2

Table 1b.04.02 Measure Rate by Hispanic Ethnicity displays the number and measure rate for patients in the following categories: Hispanic or Latino (35.8%), Not Hispanic or Latino (23.9%), Missing (0.0%). The 2020 data represents 993 discharges in the measure population from 15 hospitals.

Table 1b.04.03 Measure Rate by Race

race	rate	N
American Indian or Alaska Native	0.0%	2
Asian	18.9%	74
Black or African American	32.4%	105
Native Hawaiian or Other Pacific Islander	18.2%	11
Other Race	19.4%	222
White	30.8%	577
Missing	50.0%	2

Table 1b.04.03 Measure Rate by Race displays the number and measure rate for patients in the following categories: American Indian or Alaska Native (0.0%), Asian (18.9%), Black or African American (32.4%), Native Hawaiian or Other Pacific Islander (18.2%), Other Race (19.4%), White (30.8%). The 2020 data represents 993 discharges in the measure population from 15 hospitals.

Table 1b.04.04 Measure rate by Payer

Payer	Rate	N
Commercial	23.5%	600
Medicaid/Medicare	20.3%	237
Other	53.2%	156

Table 1b.04.04 Measure rate by Payer displays the number and measure rate for patients in the following categories: Commercial (23.5%), Medicaid/Medicare (20.3%), Other (53.2%). The 2020 data represents 993 discharges in the measure population from 15 hospitals.

[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]
See 1b.04
Okwandu, I. C., Anderson, M., Postlethwaite, D., Shirazi, A., & Torrente, S. (2021). Racial and Ethnic Disparities in Cesarean Delivery and Indications Among Nulliparous, Term, Singleton, Vertex Women. Journal of racial and ethnic health disparities, 10.1007/s40615-021-01057-w. Advance online publication. https://doi.org/10.1007/s40615-021-01057-w
Okwandu et al (2021) conducted a retrospective cohort study of NTSV deliveries at Kaiser Permanente Northern California from 1/1/2016 to 6/30/2017. The study included 16,587 racially/ethnically diverse women who met inclusion and exclusion criteria. Exclusion criteria included cesarean delivery indications of elective, malpresentation, or previa. To assess the likelihood of cesarean delivery by race/ethnicity multivariable logistic regression models adjusted for maternal, neonatal, and facility factors were used. Additional testing was performed to evaluate the odds of cesarean for the indications of failure to progress and fetal intolerance by race/ethnicity. The results of the adjusted logistic regression models showed all race and ethnic categories had higher odds of cesarean deliveries compared to White women. In

comparison with White women, Black women had greater odds of fetal intolerance as an indication, while Hispanic and Asian women had greater odds of failure to progress. Observed disparities in cesarean delivery rates were not explained by maternal, neonate, and facility factors.

Table 3 Adjusted odds ratios of cesarean delivery, fetal intolerance of labor, and failure to progress by race/ethnicity controlling for demographic and clinical characteristics

Race/ethnicity (Reference: White)	Cesarean Delivery (n = 16,587) aOR† (95% CI‡)	Fetal intolerance of labor* (n = 3727) aOR (95% CI)	Failure to Progress* (n = 3727) aOR (95% CI)
Asian/Hawaiian/Pacific Islander	1.59 (1.44–1.76)	0.90 (0.76–1.07)	1.46 (1.22–1.74)
Black	1.73 (1.45–2.06)	1.51 (1.10–2.07)	0.77 (0.56–1.04)
Hispanic	1.43 (1.28–1.59)	0.87 (0.72–1.05)	1.25 (1.03–1.52)
Multiple races/American Indian/Alaskan Native	1.45 (1.17–1.80)	1.27 (0.86–1.86)	1.03 (0.70–1.51)

*Among women who had cesarean deliveries, 486 (13.02%) of the women who had cesarean deliveries had both fetal intolerance of labor and failure to progress as indications

†Adjusted odds ratio from logistic regression models adjusted for maternal age at delivery, income, education, marital status, obesity, gestational diabetes, gestational or chronic hypertension or preeclampsia, induction of labor, availability of midwifery services, gestational age, and neonate birth weight

‡ Confidence interval

Table 3 Adjusted odds ratios of cesarean delivery, fetal intolerance of labor, and failure to progress by race/ethnicity controlling for demographic and clinical characteristics.

Debbink, M. P., Ugwu, L. G., Grobman, W. A., Reddy, U. M., Tita, A., El-Sayed, Y. Y., Wapner, R. J., Rouse, D. J., Saade, G. R., Thorp, J. M., Jr, Chauhan, S. P., Costantine, M. M., Chien, E. K., Casey, B. M., Srinivas, S. K., Swamy, G. K., Simhan, H. N., & Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) Maternal-Fetal Medicine Units (MFMU) Network (2022). Racial and Ethnic Inequities in Cesarean Birth and Maternal Morbidity in a Low-Risk, Nulliparous Cohort. Obstetrics and gynecology, 139(1), 73–82.
https://doi.org/10.1097/AOG.0000000000004620

Debbink et. al. (2022) conducted a secondary analysis of ARRIVE, a multicenter randomized trial of induction of labor compared with expectant management at term conducted at 41 centers across the United States. The analysis was performed to evaluate differences in cesarean birth and maternal morbidity in low-risk nulliparous people at term. 1,158 of the 5,759 participants delivered by cesarean which included 1,404 (24.3%) participants who identified as non-Hispanic Black, 1,670 (29.0%) as Hispanic, and 2,685 (46.6%) as non-Hispanic White. When compared with non-Hispanic White people, an increased relative risk of cesarean birth was found among non-Hispanic Black (adjusted relative risk [aRR] 1.21, 95% CI 1.03–1.42) and Hispanic (aRR 1.26, 95% CI 1.08–1.46) people. Excess maternal morbidity attributed to cesarean birth among non-Hispanic Black and Hispanic people was an estimated 15.8% (95%CI 2.1–48.7%) and 16.5% (95% CI 4.0–44.0%) respectively. Low-risk, term, nulliparous non-Hispanic Black and Hispanic people delivery by cesarean more frequently than low-risk, term, nulliparous non-Hispanic White people. These findings are important as they suggest the safe reduction of the primary cesarean birth rate among Black and Hispanic people may help to address equity in surgical outcomes and improve inequities in maternal morbidity.

[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-02 Cesarean Birth
[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]
This measure assesses the number of nulliparous women with a term, singleton baby in a vertex position delivered by cesarean birth.
[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
[Response Ends]

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

[Response Begins]
Safety: Overuse
[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]
https://www.jointcommission.org/measurement/specification-manuals/electronic-clinical-quality-measures/
Please refer to 2021 Reporting Period specifications.

[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: 0471e_PC02_eCQMFlow2020.pdf
Attachment: 0471e_ePC02_CesareanBirth_v2.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: 0471e_ePC02_eCQM_Value Sets_2021_ReportingYear.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 who deliver by cesarean section.
[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. Cesarean birth is represented with the QDM datatype and value set of "Procedure, Performed: Cesarean Birth (OID:2.16.840.1.113883.3.117.1.7.1.282). The delivery procedure must be performed during the encounter.
2. The measure looks to see if the Cesarean birth was performed during the inpatient encounter.
3. To access the value sets for the measure, please visit the Value Set Authority Center (VSAC), sponsored by the National Library of Medicine, at https://vsac.nlm.nih.gov/. A list of value sets for the measure is attached in the Excel workbook provided for question sp.11.

[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]
Inpatient hospitalizations for nulliparous patients delivered of a live term singleton newborn >= 37 weeks gestation.
[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]

1. Nulliparous patients are represented by the following QDM datatypes and value sets

Assessment, Performed: Parity (Result = 0) using Parity LOINC Direct Reference Code 11977-6
OR
Assessment, Performed: Gravida (Result =1) using Gravida (# Pregnancies) LOINC Direct Reference Code 11996-6)
OR
Assessment, Performed: Preterm (result = 0) AND Assessment, Performed: Term Newborn (result = 0) using Preterm LOINC Code 11637-6AND Term Newborn LOINC Direct Reference Code 11639-2
The nulliparous conditions must be resulted <= 42 weeks before the time of delivery. The relevant date/time (when the assessment was actually performed) of the nulliparous condition is used. Time of Delivery is represented by the QDM datatype of Assessment, Performed: Date and time of obstetric delivery using Date and time of obstetric delivery LOINC Direct Reference Code 93857-1.

2. The logic determines gestational age as follows:
   a. For the Estimated Due Date (EDD), the QDM datatype and value set of 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 one day or less before or on the delivery date/time.
   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.
   c. The logic includes a function which calculates the gestational age. This function reflects the ACOG ReVITALize Guidelines for Calculated 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.

3. If the necessary data elements are not available to calculate CGA, CGA will be null. Then the estimated gestational age which is derived from the QDM datatype and value set of Assessment, Performed: Estimated Gestational Age at Delivery using SNOMEDCT Value Set (2.16.840.1.113762.1.4.1045.26 is used.

4. Live singleton newborns are represented by the QDM datatype Encounter Performed, Diagnosis: Delivery of Singleton using ICD10 and SNOMED codes (2.16.840.1.113762.1.4.1045.99)

To access the value sets for the measure, please visit the Value Set Authority Center (VSAC), sponsored by the National Library of Medicine, at https://vsac.nlm.nih.gov/. A list of value sets for the measure is attached in the Excel workbook provided for question sp.11.

[Response Ends]

sp.16. Describe the denominator exclusions.

Brief narrative description of exclusions from the target population.

[Response Begins]
Inpatient hospitalizations for patients with abnormal presentation or placenta previa during the encounter.
[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]

1. Encounter Performed, Diagnosis: Abnormal Presentation (2.16.840.1.113762.1.4.1045.105) and Assessment Performed: Abnormal Presentation (2.16.840.1.113762.1.4.1045.105) are used to identify patients with Abnormal Presentation for exclusion from the denominator.
2. Encounter Performed, Diagnosis: Placenta Previa (2.16.840.1.113762.1.4.1110.37) is used to identify patients with Placenta Previa for exclusion from the denominator.

[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]
Not applicable; this measure is not stratified.
[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]
No risk adjustment or risk stratification
[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]
Not applicable; this measure does not use a sample.
[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]
Hospitals report EHR data using Certified Electronic Health Record Technology (CEHRT), and by submitting Quality Reporting Document Architecture Category 1 (QRDA-1).
[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]
07-01-2020 – 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]

In 2020, The Joint Commission introduced the Cesarean Birth measure (ePC02) as one of the available eCQMs hospitals could choose for data submission to meet ORYX requirements. 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 6 sites consisting of 15 hospitals submitted production data for one quarter of calendar year 2020. These data were used for all of the testing provided with the exception of validity testing, which used a subset of the six sites. TJC reached out to all 15 hospitals to recruit sites willing to participate in validity testing on the data submitted. Two pilot sites (7 hospitals) volunteered. One site is a system representing 6 hospitals where the Epic system is used. The 7th hospital is a stand-alone facility that uses Meditech.

In order to capture the individual hospital level characteristics, we referenced the American Hospital Association (AHA) DataQuery ™ product, at the URL https://guide.prod.iam.aha.org/dataquery/reports, accessed November 11, 2021. Table 2a.05.01 is a summary of the hospital characteristics as reported to AHA DataQuery™.

Table 2a.05.01 Hospital Characteristics

Hospital Characteristics	Values
State Represented	6 states represented and Puerto Rico CA, CT, IA, MI, OH, PR, WA
Hospital Location	12 Urban 3 Rural
Control / Ownership Type	9 Nongovernment (not-for-profit) 3 Government 3 For Profit
Primary Service	14 General medical and surgical 1 Specialty
Teaching Affiliation	11 Minor teaching 4 Non-teaching
System member	10 Yes 5 No
Beds (total facility)	6 <100 beds 5 100-399 beds 4 400 + beds
Total births (per year)	Range 165-5323 6 100-499 4 500-999 4 1000-4999 1 Greater than 5000

Table 2a.05.01 Hospital Characteristics describes the state represented, hospital location, control/ownership type, primary service, teaching affiliation, system member, beds and total births for the 15 hospitals that submitted data for 2020 discharges.

[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 Pilot Hospitals (Note: All records received from hospitals are used in this analysis to provide a demographic profile of all patients, not only those that fell into the measure).

Category	1.1	1.2	1.3	1.4	1.5	1.6	2	3.1	3.2	3.3	4	5	6.1	6.2	6.3	Across Hospitals
Maternal Age in Years	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
0-18	0 (0)	0 (0)	2 (0.7)	1 (1.9)	1 (1.1)	3 (3.4)	0 (0)	5 (2)	0 (0)	0 (0)	0 (0)	0 (0)	14 (1.2)	2 (1.4)	1 (0.5)	29 (1)
18-25	24 (28.6)	17 (43.6)	84 (31.3)	21 (39.6)	25 (26.3)	31 (35.2)	23 (28)	105 (42.7)	41 (38)	18 (64.3)	31 (34.1)	36 (14.6)	184 (16.2)	31 (22)	19 (9.4)	690 (23.7)
25-30	31 (36.9)	11 (28.2)	68 (25.4)	13 (24.5)	36 (37.9)	24 (27.3)	29 (35.4)	81 (32.9)	32 (29.6)	7 (25)	33 (36.3)	57 (23.1)	222 (19.6)	39 (27.7)	35 (17.2)	718 (24.7)
30-35	22 (26.2)	5 (12.8)	75 (28)	12 (22.6)	23 (24.2)	22 (25)	18 (22)	39 (15.9)	24 (22.2)	2 (7.1)	22 (24.2)	93 (37.7)	382 (33.7)	40 (28.4)	78 (38.4)	857 (29.5)
35-40	7 (8.3)	3 (7.7)	33 (12.3)	6 (11.3)	5 (5.3)	6 (6.8)	9 (11)	11 (4.5)	9 (8.3)	1 (3.6)	4 (4.4)	49 (19.8)	263 (23.2)	24 (17)	50 (24.6)	480 (16.5)
40-45	0 (0)	2 (5.1)	5 (1.9)	0 (0)	4 (4.2)	2 (2.3)	3 (3.7)	5 (2)	2 (1.9)	0 (0)	1 (1.1)	12 (4.9)	65 (5.7)	5 (3.5)	20 (9.9)	126 (4.3)
45-50	0 (0)	1 (2.6)	1 (0.4)	0 (0)	1 (1.1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	4 (0.4)	0 (0)	0 (0)	7 (0.2)
50+	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.1)	0 (0)	0 (0)	1 (0)
Race	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
American Indian or Alaska Native	0 (0)	0 (0)	0 (0)	0 (0)	1 (1.1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (1.6)	0 (0)	4 (0.4)	2 (1.4)	1 (0.5)	9 (0.3)
Asian	1 (1.2)	5 (10.4)	3 (1.1)	0 (0)	1 (1.1)	0 (0)	5 (6.1)	0 (0)	0 (0)	0 (0)	3 (4.7)	5 (2)	156 (13.7)	14 (9.9)	7 (3.4)	200 (6.9)
Black or African American	3 (3.6)	0 (0)	60 (22.4)	1 (1.9)	4 (4.2)	14 (15.9)	26 (31.7)	8 (3.3)	9 (8.6)	2 (7.1)	5 (7.8)	18 (7.3)	153 (13.5)	23 (16.3)	0 (0)	326 (11.3)
Native Hawaiian or Other Pacific Islander	0 (0)	28 (58.3)	1 (0.4)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	28 (2.5)	5 (3.5)	4 (2)	66 (2.3)
White	77 (91.7)	6 (12.5)	185 (69)	46 (86.8)	82 (86.3)	67 (76.1)	44 (53.7)	233 (96.7)	96 (91.4)	26 (92.9)	55 (85.9)	174 (70.4)	301 (26.5)	51 (36.2)	122 (60.1)	1565 (54.3)

Category	1.1	1.2	1.3	1.4	1.5	1.6	2	3.1	3.2	3.3	4	5	6.1	6.2	6.3	Across Hospitals
Other Race	3 (3.6)	9 (18.8)	19 (7.1)	6 (11.3)	7 (7.4)	7 (8)	7 (8.5)	0 (0)	0 (0)	0 (0)	0 (0)	50 (20.2)	493 (43.4)	46 (32.6)	69 (34)	716 (24.8)
Ethnicity	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Hispanic or Latino	3 (3.6)	30 (90.9)	18 (6.7)	8 (15.1)	4 (4.2)	13 (14.8)	6 (7.3)	239 (98.8)	105 (97.2)	28 (100)	4 (5.3)	37 (15)	413 (36.4)	57 (40.4)	85 (41.9)	1050 (36.4)
Non-Hispanic or Latino	81 (96.4)	3 (9.1)	250 (93.3)	45 (84.9)	91 (95.8)	75 (85.2)	76 (92.7)	3 (1.2)	3 (2.8)	0 (0)	72 (94.7)	210 (85)	722 (63.6)	84 (59.6)	118 (58.1)	1833 (63.6)
Primary Payer	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
BC Managed Care -- Other	12 (12.4)	2 (4.9)	28 (8.1)	17 (24.6)	31 (28.7)	11 (10.9)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	101 (3.3)
BC Managed Care -- PPO	3 (3.1)	0 (0)	26 (7.5)	2 (2.9)	4 (3.7)	4 (4)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	39 (1.3)
BLUE CROSS/BLUE SHIELD	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	16 (16.7)	0 (0)	0 (0)	0 (0)	1 (1.1)	51 (20)	0 (0)	0 (0)	0 (0)	68 (2.2)
Charity	1 (1)	1 (2.4)	0 (0)	0 (0)	0 (0)	0 (0)	1 (1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	3 (0.1)
Commercial Managed Care - HMO	7 (7.2)	5 (12.2)	37 (10.7)	1 (1.4)	4 (3.7)	7 (6.9)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	61 (2)
Commercial Managed Care - POS	4 (4.1)	5 (12.2)	13 (3.8)	0 (0)	2 (1.9)	1 (1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	25 (0.8)
Commercial Managed Care - PPO	11 (11.3)	3 (7.3)	31 (9)	0 (0)	0 (0)	10 (9.9)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	55 (1.8)
Managed Care (private)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	86 (33.7)	0 (0)	0 (0)	0 (0)	86 (2.8)
Medicaid	1 (1)	4 (9.8)	4 (1.2)	10 (14.5)	9 (8.3)	1 (1)	6 (6.3)	0 (0)	0 (0)	0 (0)	0 (0)	98 (38.4)	280 (24.3)	74 (50.3)	50 (22.5)	537 (17.3)
Medicaid HMO	38 (39.2)	0 (0)	161 (46.5)	20 (29)	39 (36.1)	56 (55.4)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	314 (10.1)

Category	1.1	1.2	1.3	1.4	1.5	1.6	2	3.1	3.2	3.3	4	5	6.1	6.2	6.3	Across Hospitals
Medicare	1 (1)	0 (0)	1 (0.3)	0 (0)	1 (0.9)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.4)	3 (0.3)	1 (0.7)	0 (0)	8 (0.3)
No Typology Code available for payment source	10 (10.3)	0 (0)	11 (3.2)	5 (7.2)	10 (9.3)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.1)	0 (0)	1 (0.5)	38 (1.2)
Other Managed Care	7 (7.2)	0 (0)	26 (7.5)	0 (0)	5 (4.6)	9 (8.9)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	47 (1.5)
Private Health Insurance	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	14 (14.6)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	846 (73.4)	69 (46.9)	164 (73.9)	1093 (35.2)
Self-pay	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	12 (12.5)	0 (0)	0 (0)	0 (0)	2 (2.2)	1 (0.4)	0 (0)	0 (0)	0 (0)	15 (0.5)
TRICARE (CHAMPUS)	1 (1)	0 (0)	2 (0.6)	1 (1.4)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	88 (96.7)	2 (0.8)	0 (0)	0 (0)	0 (0)	94 (3)
Unavailable	1 (1)	21 (51.2)	6 (1.7)	13 (18.8)	3 (2.8)	2 (2)	47 (49)	246 (100)	108 (100)	28 (100)	0 (0)	16 (6.3)	22 (1.9)	3 (2)	7 (3.2)	523 (16.8)

Table 2a.06.01 Patient Characteristics for Pilot Hospitals displays the number of cases and percentage for Maternal Age in Years, Race, Ethnicity and Primary Payer.
* Cell intentionally left empty
[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]
Rationale and data provided under risk adjustment section is from the 0471 submission.
[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]
No patient-level sociodemographic variables are used in the measure. There was considerable variability in the distribution of patient socio-demographic characteristics across hospitals, but we did not analyze differences in measure rates over these variables due to the relatively small sample size.
[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)
[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]
See validity testing section of data elements.
[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]
N/A
[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]
N/A
[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)
[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 2 pilot sites (7 hospitals). This includes 1 system of 6 hospitals and one stand-alone hospital. We reviewed 34 charts at the stand-alone hospital and 15 charts for each hospital in the system for a total of 90 charts, and a combined total for the 2 pilot sites of 124 charts. It should be noted that one chart was removed from the study as it did not have a delivery procedure on the encounter, thereby leaving a total of 123 charts in the validity study. The review included two different EHR vendors (Epic and Meditech).
Sample size calculations using a kappa of 0.60, 80% power, alpha of 5%, and a confidence interval width of 0.20 gave a required sample size of 101 for the validity study. this is the content field where there isn't a ton of stuff in the editor. You can see well maybe not that well...kljasfdkjlfdsakjlnm

Due to COVID-19, onsite validity testing visits were transitioned to a virtual visit approach. Validity visits were conducted during October and November of 2021 by The Joint Commission staff with the support of a hospital site abstractor. The purpose of the visits was to assess measure 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.
1. 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 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.
2. Analysis
Testing methodology is outlined below:
a. All clinical data elements and all editable demographic elements are scored.
b. All measure data are re-abstracted with original data having been blinded so that the re-abstraction is not biased.
c. Re-abstracted data are compared with original data for each data element to identify missing or erroneous data. Ideally, data element agreement rates should exceed 80%.
d. Overall performance measure outcome rates were calculated on all cases submitted by each 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 0.60 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

[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 for the 2 pilot sites (7 hospitals) This includes 1 system of 6 hospitals and one stand-alone hospital. We reviewed 34 charts at the stand-alone hospital and 15 charts for each hospital in the system for a total of 90 charts, and a combined total for the 2 pilot sites of 124 charts. It should be noted that one chart was removed from the study as it did not have a delivery procedure on the encounter, thereby leaving a total of 123 charts in the validity study. The review included two different EHR vendors (Epic and Meditech).
Sample size calculations using a kappa of 0.60, 80% power, alpha of 5%, and a confidence interval width of 0.20 gave a required sample size of 101 for the validity study. Table 2b.03.01 displays the PPV (agreement rate) for the numerator among delivery encounters clinically adjudicated in validity testing. The PPV rate was 94% at Pilot Site 1 and 0% at Pilot Site 2. During validity testing, Pilot Site 2 identified a mitigation plan for future data submissions.

Table 2b.03.01. Agreement Statistics for Measure Numerator between EHR Extraction and Manual Chart Abstraction (PPV) (Validity Testing, 2 Test Sites, 7 hospitals)

Pilot Sites	# Of Numerator Events Verified by Validity Testing	# Of Numerator Events from EHR	Positive Predictive Value (PPV)
Pilot Site 1	32	30	94%
Pilot Site 2	6	0	0%
Across 2 Pilot Sites (7 hospitals)	38	30	79%

Agreement Statistics for Measure Numerator between EHR Extraction and Manual Chart Abstraction (PPV) (Validity Testing, 2 Test Sites, 7 hospitals)

Table 2b.03.02 displays the sensitivity, specificity, and negative predictive value (NPV). Specificity is high for both sites and sensitivity is high for Site 1 but low for Site 2. This means that the probability of the EHR data detecting a true cesarean section during a delivery hospitalization based on the abstracted data ('gold standard') is 87.5% for Site 1, 0% for Site 2 and 73.7% overall (sensitivity). Site 2's low sensitivity rate can be explained by cases not qualifying for the initial population as time of delivery was missing or gravida/para/term/preterm were incorrect. This hospital uses a standalone OB documentation system that does not interface completely with the electronic health record (Meditech). The OB documentation is present in Meditech in non-discrete fields in a .pdf format. During validity testing the hospital identified a mitigation plan for future data submissions.

The probability of the EHR data accurately identifying that no cesarean section occurred during a delivery hospitalization based on abstracted data was 96.5% for Site 1 and 100% for Site 2, and 97.7% overall (specificity). NPV was 93.2% and 82.4% for Sites 1 and 2 respectively, 89.3% overall, indicating the EHR data identified that a cesarean section did not occur, and the chart abstraction confirmed a cesarean section 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	87.5%	96.5%	93.2%
Pilot Site 2	0%	100%	82.4%
Across 2 Pilot Sites (7 hospitals)	73.7%	97.7%	89.3%

Measure Score Validity Statistics for Sample Between EHR Extraction and Manual Chart Abstraction (Sensitivity, Specificity, NPV)

Measure Outcome Agreement Rates:

Overall, the study revealed ePC-02 to have a good measure outcome agreement rate of 83.7% with a kappa score of 0.750 indicating substantial agreement. (See Table 2b.03.03 Measure Outcome Agreement Rates)

Table 2b.03.03 Measure Outcome Agreement Rates

Pilot Site	N	Agreement Rate	kappa
Pilot Site 1	89	89.9%	0.831
Pilot Site 2	34	67.7%	0.477
Total	123	83.7%	0.750

Measure Outcome Agreement Rates

Table 2b.03.04 Data Element Agreement Rates

Characteristic	Data Element Name	Site #1	Site #1	Site #1	Site #2	Site #2	Site #2	Total	Total	Total
*	*	Match	N	Rate	Match	N	Rate	Match	N	Rate
Demographics	DOB	89	89	100%	34	34	100%	123	123	100%
Demographics	ONC Administrative Sex Code	89	89	100%	34	34	100%	123	123	100%
Demographics	Race	88	89	99%	34	34	100%	122	123	99%
Demographics	Ethnicity	88	89	99%	34	34	100%	122	123	99%
Demographics	Payer	89	89	100%	34	34	100%	123	123	100%
Encounter	Encounter, Performed : Encounter Inpatient	21	89	24%	34	34	100%	55	123	45%

Characteristic	Data Element Name	Site #1	Site #1	Site #1	Site #2	Site #2	Site #2	Total	Total	Total
History	Admission Date Time (Relevant Period Start Time)	89	89	100%	34	34	100%	123	123	100%
History	Discharge Date Time (Relevant Period End Time)	89	89	100%	34	34	100%	123	123	100%
Dx	Abnormal Presentation Diagnosis Code	83	89	93%	34	34	100%	117	123	95%
Dx	Delivery of Singleton Diagnosis Code (ICD10)	89	89	100%	34	34	100%	123	123	100%
Dx	Delivery of Singleton Diagnosis (SNOMED)	89	89	100%	0	0		89	89	100%
Dx	Placenta Previa Diagnosis Code	89	89	100%	34	34	100%	123	123	100%
Procedures	Cesarean Section Procedure Code	39	39	100%	12	12	100%	51	51	100%
Procedures	Cesarean Section Procedure Date	39	39	100%	12	12	100%	51	51	100%
Procedures	Delivery Procedure Code	89	89	100%	34	34	100%	123	123	100%
Procedures	Delivery Procedure Date	89	89	100%	33	34	97%	122	123	99%
Delivery Details	Assessment, Performed: Date and time of obstetric delivery, Author Date Time[1]	89	89	100%	14	34	41%	103	123	84%
Delivery Details	Assessment, Performed: Estimated Gestational Age at Delivery, Author Date Time[1]	89	89	100%	3	34	9%	92	123	75%
Delivery Details	Assessment, Performed: Estimated Gestational Age at Delivery, result[1]	89	89	100%	27	34	79%	116	123	94%
Pregnancy History	Assessment, Performed: Births.preterm - Author Date Time	89	89	100%	0	9	0%	89	98	91%
Pregnancy History	Assessment, Performed: Births.preterm – Result	89	89	100%	0	12	0%	89	101	88%
Pregnancy History	Assessment, Performed: Births.term - Author Date Time	89	89	100%	0	15	0%	89	104	86%
Pregnancy History	Assessment, Performed: Births.term – Result	86	89	97%	0	17	0%	86	106	81%
Pregnancy History	Assessment, Performed: Parity - Author Date Time	89	89	100%	15	34	44%	104	123	85%

Characteristic	Data Element Name	Site #1	Site #1	Site #1	Site #2	Site #2	Site #2	Total	Total	Total
Pregnancy History	Assess Perf Parity - Result	88	89	99%	25	34	74%	113	123	92%
Pregnancy History	Assessment, Performed: pregnancies (gravida) - Author Date Time	89	89	100%	17	34	50%	106	123	86%
Pregnancy History	Assessment, Performed: pregnancies (gravida) - Result	89	89	100%	31	34	91%	120	123	98%
*	TOTALS	2223	2303	97%	597	757	79%	2820	3060	92%

Data Element Agreement Rates

* Cell intentionally left empty

[1]See Table 2b.03.05 Data Element Agreement Rates Shared Data Elements

It should be noted that in the latest version of ePC02, the initial population and determination of gestational age for the denominator is aligned with ePC07 (Severe Obstetrics Complication) which is a measure under development. During the development of ePC07 in 2021, validity testing was completed for 15 individual hospitals (1 system of 10 hospitals and 5 individual hospitals). Over 200 records were subjected to validity testing in 2021. The following 6 data elements in Table 2b.03.05 are used by both ePC02 and ePC07. The overall validity for these 6 data elements is high at 94.1%.

Table 2b.03.05 Data Element Agreement Rates Shared Data Elements

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
*	Match	N	Rate	Match	N	Rate	Match	N	Rate	Match	N	Rate	Match	N	Rate	Match	N	Rate	Match	N	Rate
Relevant Date Time Assessment, Performed: Date and time of obstetric delivery	35	36	97%	30	31	97%	34	35	97%	36	36	1	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	1	28	30	93%	36	36	100%	199	204	98%

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: Delivery date Estimated	34	36	94%	0	31	0%	34	35	97%	36	36	1	28	30	93%	36	36	100%	168	204	82%
Result: Delivery date Estimated	34	36	94%	30	31	97%	35	35	100%	36	36	1	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	1	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	1	30	30	100%	34	36	94%	195	204	96%
TOTALS	209	216	97%	138	185	75%	206	210	98%	216	216	1	172	180	96%	210	216	97%	1151	1223	94%

Data Element Agreement Rates Shared Data Elements

Table 2b.03.05 Data Element Agreement Rates Shared Data Elements. It should be noted that in the latest version of ePC02, the initial population and determination of gestational age for the denominator is aligned with ePC07 (Severe Obstetrics Complication) which is a measure under development. During the development of ePC07 in 2021, validity

testing was completed for 15 individual hospitals (1 system of 10 hospitals and 5 individual hospitals). Over 200 records were subjected to validity testing in 2021. The 6 data elements in Table 2b.03.05 are used by both ePC02 and ePC07. The overall validity for these 6 data elements is high at 94.1%.
* Cell intentionally left empty

[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): For Site 1, in almost all delivery encounters with a numerator event adjudicated, the delivery encounters with a cesarean section in the EHR data were shown to have a cesarean section 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. The absence of a perfect PPV does not threaten validity as we do not expect any systematic error in this small amount of disagreement across hospitals that might bias the measure results. Site 2 had poor positive predictive accuracy due to data collection issues specific to their site explained below.
Sensitivity, specificity, and negative predictive value (NPV): Specificity and sensitivity results for site 1 indicate a high probability of the EHR data detecting a true cesarean section based on the abstracted data ('gold standard'), and a high probability of the EHR data accurately identifying that cesarean section occurred during a delivery hospitalization. The strong NPV results indicate that when EHR data indicated a cesarean section did not occur, the chart abstraction confirmed that a cesarean section did not occur. Site 2 had poor positive sensitivity due to data collection issues specific to their site explained below.
Pilot Site 1: 30 numerator events were identified based on submitted data. Thirty-two numerator events were evaluated during validity testing. During validity testing, it was identified that 6 records were coded as malpresentation of fetus. However, upon clinical adjudication, documentation was found that the fetus was in the vertex presentation. Therefore, these six cases were no longer denominator exclusions. Four of the six cases became numerator cases, 2 of the cases became denominator cases. Two additional cases originally qualified as numerator cases but upon validity testing and clinical adjudication were found to not meet the denominator. The preterm/term births were submitted as zero qualifying them for the denominator based on EHR data. However, during validity testing and clinical adjudication the patients were found to have one term birth disqualifying them for the denominator. Overall, the net change in the numerator count was plus 2.
Pilot Site 2: No numerator events were identified based on submitted data. Six numerator events were evaluated during validity testing. During validity testing six cases that were originally excluded from the denominator due to a missing time of delivery were found to meet the numerator when time of delivery was provided. This hospital uses a standalone OB documentation system that does not interface completely with the electronic health record (Meditech). The OB documentation is present in Meditech in non-discrete fields in a .pdf form.
Measure Outcome Agreement Rates Analysis:
Pilot Site 1: 89 records across 6 individual hospitals exhibited an 89.9% measure outcome agreement rate with a kappa score of 0.831 indicating almost perfect agreement. Six of the nine mismatches were coded as malpresentation of fetus. However, review of the clinical record revealed that the fetus was in vertex presentation. Therefore, these six cases no longer were denominator exclusions. The remaining 3 cases were mismatches since the para or # term births were incorrect. Based on the submitted data, the patient did not qualify for the denominator. With the adjudicated data, the patient qualified for the denominator or numerator.
Pilot Site 2: 34 records for Pilot Site 2 exhibited a measure outcome agreement rate of 67.7% with kappa score of 0.477 indicating moderate agreement. In total, 11 cases mismatched. In all 11 cases, the patient did not qualify for the initial population as time of delivery was missing or gravida/para/term/preterm were incorrect. This hospital uses a standalone OB documentation system that does not interface completely with the electronic health record (Meditech). The OB documentation is present in Meditech in non-discrete fields in a .pdf format.
Data Element Agreement Rate Analysis:
Comment on feasibility scorecard in relationship to validity: As evidenced on the feasibility scorecards created for this measure, only one data element (estimated gestational age author date/time) was scored as 0 by Pilot Site 2 for data accuracy. Validity testing proved this data element to be problematic along with other data elements. The hospital has identified a mitigation plan for future data submissions.
Overall, the data element agreement rate for all sites was excellent at a score of 92.2%.

Pilot Site 1 demonstrated an excellent agreement rate of 96.5%. In 21 out of 89 cases, the admission type was erroneously mapped to emergency when it should have been elective. This had no impact on the measure outcome. As already mentioned, six cases were miscoded as malpresentation of the fetus. None of the mismatches were due to missing data.
Pilot Site 2 demonstrated a fair data element agreement rate of 78.9%. As already mentioned, this hospital uses a stand-alone OB documentation system that does not interface completely with the Meditech Electronic Health Record. A .pdf report is available in Meditech in non-discrete fields. Most mismatches were in the Delivery Date/Time, Estimated Gestational Age, Gravida, Para, Preterm or Term Birth fields. Of the mismatches, 57% were due to missing data in those fields.

[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]
To demonstrate meaningful differences in performance, The Joint Commission calculated a funnel plot (Spiegelhalter, 2005) for the hospital rates of the measure. In a funnel plot, the observed measure is plotted against a measure of its precision, so that the control limits form a 'funnel' around the target outcome. The 95 percent (≈2 standard deviation) and 99.8 percent (≈3 standard deviation) prediction limits are then superimposed over this plot around the overall measure rate. Those rates lying outside the confidence limits are identified as outliers.
Spiegelhalter, DJ. (2005). Funnel plots for comparing institutional performance. *Statistics in Medicine*, 24(8), 1185–1202. doi: 10.1002/sim.1970.

[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.

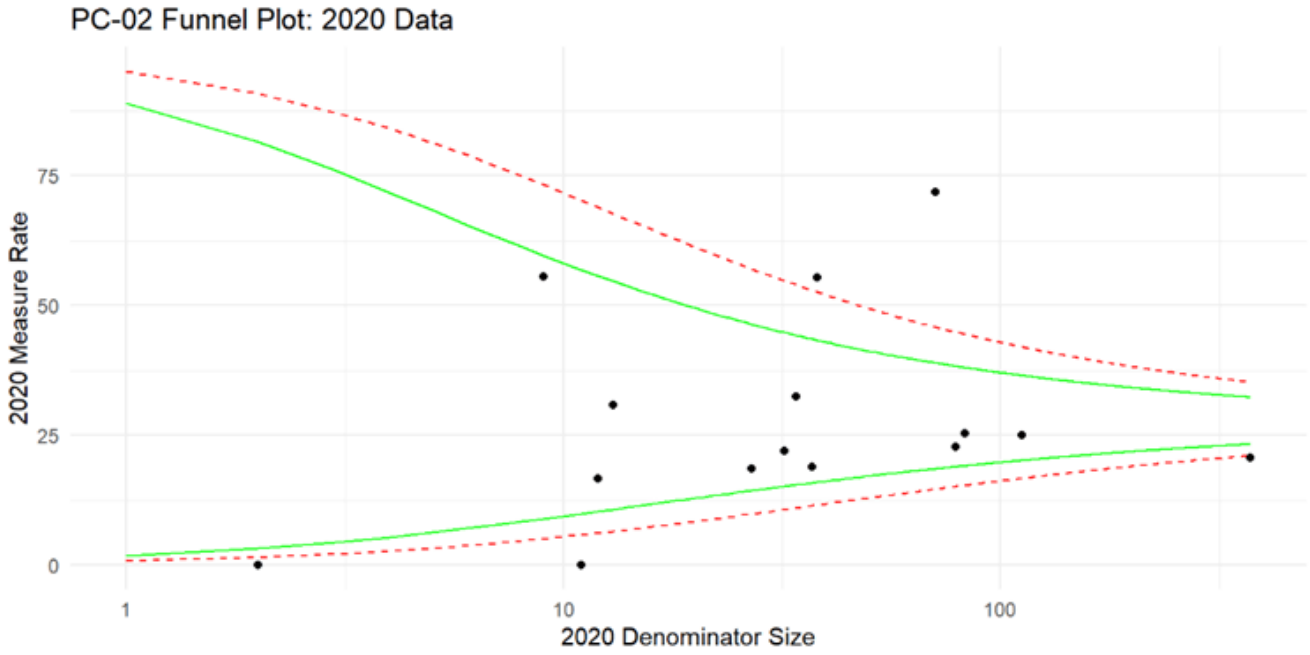
[Response Begins]
Cesarean Birth by Site
Cesarean birth rates for all 15 hospitals that submitted production data for the 2020 calendar year can be found in Table 2b.06.01 below. The first digit of the hospital ID identifies the site and the digit after the decimal indicates the hospital within that site.
Table 2b.06.01 ePC02 Cesarean Birth Rates

Hospital ID	Denominator	ePC02 Rate
1.1	34	32.4%
1.2	13	30.8%
1.3	79	22.8%
1.4	12	16.7%
1.5	32	21.9%
1.6	27	18.5%
2	11	0.0%
3.1	71	71.8%
3.2	38	55.3%
3.3	9	55.6%

Hospital ID	Denominator	ePC02 Rate
4	2	0.0%
5	112	25.0%
6.1	373	20.6%
6.2	37	18.9%
6.3	83	25.3%
Total	933	27.5%

Table 2b.06.01 ePC02 Cesarean Birth Rates displays the number of denominator cases and the Cesarean birth rate at the hospital level for the 15 hospitals that submitted 2020 discharges.
The funnel plot is displayed in Figure 2b.06.01 below. Of the 15 hospitals in the pilot, two were identified as statistically significant high outliers.

Figure 2b.06.01 ePC02 Funnel Plot: 2020 Data



Footnote: The green solid lines are the 95% confidence limits, and the dotted red lines are the 99% confidence limits.

[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]
Even with the relatively small numbers of hospitals and denominator sizes in the pilot, there were high outliers identified and there was significant variation in the measure rates. Those outliers identified with unusually low rates based on their denominators identify potential future data validation opportunities. Site 3 is located outside of the continental United States where performance is dramatically different.
[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]
As described in section 3.05, we quantitatively assessed data element feasibility during feasibility testing. Two domains on the NQF scorecard address missing data. The domain of "Data Availability" addresses the extent to which the data are readily available in a structured format. The domain of "Workflow" addresses the extent to which the data is routinely collected during clinical care. Based on feasibility testing results, the measure uses data elements that are expected to be available in structured fields of the EHR and captured as part of routine care of the patient.
As described in section 2b.03.04, we quantitatively assessed data element validity by indicating the "match" rate. A match indicates that the data submitted by the hospital matched what was reabstracted during the validation visit. In other words, a match indicates the data was not missing and was accurate. Overall, the data element agreement rate for all sites was excellent at a score of 92.2%. For the missing data, we determined the percent of mismatches for each data element that were due to missing data, separately for the two sites Table 2b.09.01 provides the frequency of missing data, and section 2b.10 provides the interpretation of the results.

[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]
In section 2b.03, Table 2b.03.04 displays the Data Element Agreement Rate. In this section we add the column "Due to Missing Data" to Table 2b.09.01 to provide the frequency of missing data.
Table 2b.09.01 Match Rate by Data Element Due to Missing Data

*	Pilot Site #1	Pilot Site #1	Pilot Site #1	Pilot Site #2	Pilot Site #2	Pilot Site #2	Pilot Site #2	Total	Total	Total	Total
Data Element Name	**Match Rate**	**N**	**Mismatch**	**Due to missing data**	**Match Rate**	**N**	**Mismatch**	**Due to missing data**	**Match Rate**	**N**	**Mismatch**	**Due to missing data**
DOB	100.0%	89	*	*	100.0%	34	*	*	100.0%	123	*	*
ONC Administrative Sex Code	100.0%	89	*	*	100.0%	34	*	*	100.0%	123	*	*
Race	98.9%	89	1	0	100.0%	34	*	*	99.2%	123	1	0
Ethnicity	98.9%	89	1	0	100.0%	34	*	*	99.2%	123	1	0
Payer	100.0%	89	*	*	100.0%	34	*	*	100.0%	123	*	*
Encounter, Performed : Encounter Inpatient	23.6%	89	68	0	100.0%	34	*	*	44.7%	123	68	0
Admission Date Time (Relevant Period Start Time)	100.0%	89	*	*	100.0%	34	*	*	100.0%	123	*	*

*	Pilot Site #1	Pilot Site #1	Pilot Site #1	Pilot Site #1	Pilot Site #2	Pilot Site #2	Pilot Site #2	Pilot Site #2	Total	Total	Total	Total
Discharge Date Time (Relevant Period End Time)	100.0%	89	*	*	100.0%	34	*	*	100.0%	123	*	*
Abnormal Presentation Diagnosis Code	93.3%	89	6	0	100.0%	34	*	*	95.1%	123	6	0
Delivery of Singleton Diagnosis Code (ICD10)	100.0%	89	*	*	100.0%	34	*	*	100.0%	123	*	*
Delivery of Singleton Diagnosis (SNOMED)	100.0%	89	*	*		0	*	*	100.0%	89	*	*
Placenta Previa Diagnosis Code	100.0%	89	*	*	100.0%	34	*	*	100.0%	123	*	*
Cesarean Section Procedure Code	100.0%	39	*	*	100.0%	12	*	*	100.0%	51	*	*
Cesarean Section Procedure Date	100.0%	39	*	*	100.0%	12	*	*	100.0%	51	*	*
Delivery Procedure Code	100.0%	89	*	*	100.0%	34	*	*	100.0%	123	*	*
Delivery Procedure Date	100.0%	89	*	*	97.1%	34	1	0	99.2%	123	1	0
Assessment, Performed: Date and time of obstetric delivery, Author Date Time	100.0%	89	*	*	41.2%	34	20	15	83.7%	123	20	15
Assessment, Performed: Estimated Gestational Age at Delivery, Author Date Time	100.0%	89	*	*	8.8%	34	31	6	74.8%	123	31	6
Assessment, Performed: Estimated Gestational Age at Delivery, result	100.0%	89	*	*	79.4%	34	7	6	94.3%	123	7	6
Assessment, Performed: Births. preterm - Author Date Time	100.0%	89	*	*	0.0%	9	9	9	90.8%	98	9	9
Assessment, Performed: Births. preterm - Result	100.0%	89	*	*	0.0%	12	12	12	88.1%	101	12	12
Assessment, Performed: Births.term - Author Date Time	100.0%	89	*	*	0.0%	15	15	15	85.6%	104	15	15
Assessment, Performed: Births.term - Result	96.6%	89	3	0	0.0%	17	17	17	81.1%	106	20	17

*	Pilot Site #1	Pilot Site #1	Pilot Site #1	Pilot Site #1	Pilot Site #2	Pilot Site #2	Pilot Site #2	Pilot Site #2	Total	Total	Total	Total
Assessment, Performed: Parity - Author Date Time	100.0%	89	*	*	44.1%	34	19	3	84.6%	123	19	3
Assess Perf Parity - Result	98.9%	89	1	0	73.5%	34	9	3	91.9%	123	10	3
Assessment, Performed: pregnancies (gravida) - Author Date Time	100.0%	89	*	*	50.0%	34	17	3	86.2%	123	17	3
Assessment, Performed: pregnancies (gravida) - Result	100.0%	89	*	*	91.2%	34	3	3	97.6%	123	3	3
TOTALS	96.5%	2303	80	0	78.9%	757	160	92	92.2%	3060	240	92

*This cell intentionally left empty.

Table 2b.09.01 Match Rate by Data Element Due to Missing Data is displayed for 27 data elements for the 2 Test Sites. Pilot Site 1 had no mismatches due to missing data while Pilot Site 2 had 92 mismatches due to missing data out of a total of 160 mismatches.

[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]

A match indicates that the data submitted by the hospital matched what was reabstracted during the validation visit. In other words, a match indicates the data was not missing and was accurate. As evidenced by the results above, there is variation between Pilot Site 1 and 2 with results of 96.5% and 78.9% respectively. Section 2b.04 outlines the root cause for the missing data for Pilot Site 2 and mitigation plans which would improve upon the number of missing data elements for this site. As already mentioned, Pilot Site 2 uses a stand-alone OB documentation system that does not interface completely with the Meditech Electronic Health Record. A .pdf report is available in Meditech in non-discrete fields. Most mismatches were in the Delivery Date/Time, Estimated Gestational Age, Gravida, Para, Preterm or Term Birth fields. Of the mismatches, 57% were due to missing data in those fields. In comparison, Pilot Site 1 had no mismatches due to missing data.

[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 site before and after the exclusions. The performance scores were re-calculated and checked for any significant change after exclusions. Since the number of 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.17.01 Frequency Distribution of ePC02 eCQM Exclusions

*	Denominator	Denominator	Denominator	Denominator	Denominator	Denominator
*	Delivery Encounters	Denominator Exclusions (Placenta Previa & Abnormal Presentation)	Denominator Exclusions Placenta Previa & Abnormal Presentation)	Denominator less Exclusions Placenta Previa & Abnormal Presentation)	Rate With Placenta Previa & Abnormal Presentation Cases Excluded	Rate with Placenta Previa & Abnormal Presentation Cases Included
Hospital Number	*	N	%	Denominator	%	%
1.1	35	1	2.9%	34	32%	34%
1.2	13	0	0.0%	13	31%	31%
1.3	86	7	8.1%	79	23%	29%
1.4	12	0	0.0%	12	17%	17%
1.5	38	6	15.8%	32	22%	34%
1.6	32	5	15.6%	27	19%	31%
2	11	0	0.0%	11	0%	0%
3.1	71	0	0.0%	71	72%	72%
3.2	38	0	0.0%	38	55%	55%
3.3	9	0	0.0%	9	56%	56%
4	2	0	0.0%	2	0%	0%
5	122	10	8.2%	112	25%	31%
6.1	399	26	6.5%	373	21%	26%
6.2	41	4	9.8%	37	19%	27%
6.3	88	5	5.7%	83	25%	30%
Total	997	64	6.4%	933	28%	32%

Table 2b.17.01 Frequency Distribution of ePC02 eCQM Exclusions displays the total number of denominator exclusions for the 15 hospitals. The denominator less exclusions and the rate with placenta previa & abnormal presentation cases excluded as well as the rate with the placenta previa and abnormal presentation cases included are displayed.

* Cell intentionally left empty

[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]
Exclusions can have an appreciable impact on measure rates; without excluding these cases measure rates increase overall by 17%, or 4.7 percentage points. Exclusion rates ranged from 0%-16%, indicating variability over sites.
Based on our analysis, exclusions occur with sufficient prevalence to warrant inclusion and maintain consistency of intent with the original chart-abstracted measure. The overall percentage of patients excluded from the denominator was 6.4 percent across all hospitals in the sample. Note that both measure exclusions are supported by clinical evidence, and they continue to be used in the recently endorsed chart-abstracted PC-02 measure. Thus, including these denominator exclusions in the measure increases the validity of the measure.

[Response Ends]

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

[Response Begins]
No risk adjustment or stratification
[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]
[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]
Rationale - thisiisiisiiisiiisiisiisiiisiis - This is a little bit faster. Here we are adding content. And this is the issue. IF we keep adding content then this will actiually go into view mode. I am not sure when theis will happen. I don't this it will happen if we use copy and paste. But I was able to see this earlier.
This measure is not risk-adjusted. When constructing the measure, the exclusion criteria were chosen to ensure that the target population would be women with nulliparous, term, singleton, vertex (NTSV) pregnancies. Nulliparous women are those experiencing their first birth. These women have a lower risk of maternal morbidity and mortality during a vaginal birth delivery than do women who have undergone a previous C-section (American College of Obstetricians and Gynecologists [ACOG], 2014). "Term" indicates a newborn at greater than or equal to 37 weeks gestation completed, which has better outcomes than a preterm birth. A "singleton" refers to the birth of a single newborn during the delivery encounter. Vertex presentations, which are those where the fetus is positioned headfirst, carry less risk than breach or transverse presentations (ACOG, 2014). The population of women in the denominator as a result of the exclusions, allow the measure to focus on a more homogeneous group of women where the greatest improvement opportunity exists as evidenced by the variation in rates of NTSV cesarean births indicating clinical practice patterns may affect this rate (ACOG, 2014). Lowering the C-section rate in NTSV pregnancies is important because C-sections may carry a higher risk of subsequent miscarriage, placental abnormalities, and repeat C-section (Keag et al., 2018). The rates of ruptured uteruses,

unplanned hysterectomies, and intensive care unit (ICU) admission are higher among women who deliver via C-section for the first time than those who deliver vaginally for the first time across all races and ethnicities. However, non-Hispanic Black women who deliver via C-section for the first time had the highest rates of uterine rupture and ICU admission compared with all other races (Centers for Disease Control and Prevention, 2015). Focusing on the NTSV population only and not excluding for other maternal medical conditions aligns with the measure intent to have a significant effect on cesarean birth rates and will encourage a decrease in C-section rates in the NTSV population which will in turn have a meaningful impact on future pregnancies and maternal health. Including a comprehensive set of maternal medical exclusions would add data collection burdens without commensurate benefit. Evidence continues to support no further risk adjustment is indicated as described below by Dr. Elliott Main, in an article that is in the review process for publication.

1. American College of Obstetricians and Gynecologists (College), Society for Maternal-Fetal Medicine, Caughey, A. B., Cahill, A. G., Guise, J. M., & Rouse, D. J. (2014). Safe prevention of the primary cesarean delivery. *American Journal of Obstetrics and Gynecology*, 210(3), 179–193. https://doi.org/10.1016/j.ajog.2014.01.026
2. Keag, O.E., Norman, J.E. & Stock, S.J. (2018). Long-term risks and benefits associated with cesarean delivery for mother, baby, and subsequent pregnancies: Systematic review and meta-analysis. *Plos Medicine*, 15(1), e1002494. doi: 10.1371/journal.pmed.1002494. eCollection 2018 Jan.
3. Centers for Disease Control and Prevention. (2015, May 20). National Vital Statistics Reports, Volume 64, Number 4 https:/www.cdc.gov/nchs/data/nvsr/nvsr64/nvsr64_04.pdf

CMQCC Analysis of SMFM Proposed Additional Diagnoses to NTSV Exclusion Code Set

We developed Nulliparous Term Singleton Vertex (NTSV) as the best cesarean measure to achieve two goals. First, we wanted to concentrate on the obstetric population at higher risk for cesarean birth, those on their first labor and birth (in contrast, multiparas who have had a vaginal birth have very low cesarean rates). And secondly to exclude common cesarean indications whose frequency varies significant among hospitals--breech, multiple gestations, and prematurity. This measure was adopted by Heathy People 2010, 2020 and 2030 and has been reported annually for every state. Unfortunately, the National Center for Health Statistics simplified the name for public consumption as "low-risk first-birth" cesarean rate. The measure was never intended to exclude all high-risk conditions which may affect the cesarean rate but to account for those that could have a significant effect and were maldistributed in a meaningful way.

In July 2017, the Society for Maternal-Fetal Medicine Health Policy and Advocacy Committee and Society for Maternal-Fetal Medicine Coding Committee published an expert opinion suggesting additional ICD-10-CM codes that to add to the definition of low-risk birth for the purpose of cesarean birth calculation (Armstrong et al., 2017). They specifically sought diagnosis codes that represented "clinically relevant risk factors that are absolute or relative contraindications to vaginal birth." The choice of codes was not based on actual data but solely on expert opinion.

In the analysis that follows we will sequentially walk through the frequency of these codes in hospitals of different levels of care (Table 1); the cesarean rate for these indications (within the NTSV population), again stratified by hospital level (Table 2); and lastly, revised NTSV cesarean rates should any or all of those indications be added to the exclusion list, also stratified by hospital level (Table 3).

The data indicates: (1) these diagnoses are very low frequency within the NTSV population (i.e. a large number of these cases occur either in multiparous or in preterm populations already excluded); (2) when they do occur in the NTSV population, their cesarean rate is extraordinarily high (generally <50%, certainly not the "absolute or relative contraindications to vaginal birth" as proposed by SMFM); and (3) their addition to the exclusion list leads to a minimal change in the NTSV Cesarean rate across the board-i.e. high level hospitals were not affected more than medium or lower level facilities. Therefore, we do not recommend adding these additional codes to the measure definition.

The following tables are taken from a manuscript in preparation.

Analysis of SMFM Proposed Additions to NTSV Exclusion Code Set

Base population: NTSV PC-02 population (ICD-10) in all 238 California hospitals, 2016-2017 (308,319 women giving birth)

All California hospitals were divided into 6 types: University hospital (main campus), Critical Access Hospital or by American Academy of Pediatrics Levels of Neonatal Care with Levels 3 and 4 being regional centers.

Table 2b.21.01 Frequency per 1,000 births of selected major obstetric complications (among NTSV PC-02 population)

*	Hospital Type	Hospital Type	Hospital Type	Hospital Type	Hospital Type	Hospital Type
Diagnosis Groups Proposed For Exclusion (based on ICD-10 codes)	University (main campus) (Hosp N=9) (Pt N=15,071)	AAP Level 3/4 (Hosp N=108) (Pt N=211,903)	AAP Level 2 (N=57) (Pt N=65,686)	AAP Level 1 (not Critical Access) (Hosp N=61) (Pt N=29,046)	Critical Access (Hosp N=12) (Pt N=1,684)	All Hospitals (Hosp N=238) (Pt N=308,319)
Care of Fetal anomalies	46 (3.1)	224 (1.1)	37 (0.6)	2 (0.1)	1 (0.6)	264 (0.9)
HIV	23 (1.5)	49 (0.2)	10 (0.2)	1 (0.0)	0 (0.0)	60 (0.2)
Severe Preeclampsia	46 (3.1)	552 (2.6)	171 (2.6)	79 (2.7)	8 (4.8)	810 (2.6)
Cardiovascular	332 (22.0)	1584 (7.5)	357 (5.4)	120 (4.1)	15 (8.9)	2076 (6.7)
Kidney HTN	5 (0.3)	35 (0.2)	4 (0.1)	1 (0.0)	0 (0.0)	40 (0.1)
Cerebral Thrombosis	0 (0.0)	1 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	2 (0.0)
Previa expanded	2 (0.1)	19 (0.1)	5 (0.1)	1 (0.0)	0 (0.0)	25 (0.1)
Low-lying placenta	31 (2.1)	307 (1.4)	91 (1.4)	45 (1.5)	2 (1.2)	445 (1.4)
Accreta	6 (0.4)	75 (0.4)	22 (0.3)	9 (0.3)	0 (0.0)	106 (0.3)
Abruption	3 (0.2)	21 (0.1)	4 (0.1)	2 (0.1)	0 (0.0)	27 (0.1)
Cord Prolapse	11 (0.7)	210 (1.0)	49 (0.7)	48 (1.7)	3 (1.8)	310 (1.0)
Vasa Previa	3 (0.2)	35 (0.2)	1 (0.0)	3 (0.1)	0 (0.0)	39 (0.1)
Any of the above	505 (33.5)	3078 (14.5)	745 (11.3)	305 (10.5)	29 (17.2)	4157 (13.5)

* Cell intentionally left empty

Note: there were several diagnosis groups seen more often in University hospitals than in other hospital types. However, the actual rates were still low (these are per 1,000 birth frequencies) and we went further in the next tables to examine if the cesarean rates were very high for these complications and whether excluding them would actually change the overall NTSV cesarean rates.

Table 2b.21.01 Frequency per 1,000 births of selected major obstetric complications (among NTSV PC-02 population). Five hospital type categories are displayed horizontally: University, AAP Level 3 / 4, AAP Level 2, AAP Level 1, Critical Access and Total. Vertically 12 Diagnosis Groups Proposed For Exclusion (based on ICD-10 codes) are displayed.

Table 2b.21.02 Cesarean Delivery Rate (%) for selected major obstetric complications (among NTSV PC-02 population)

*	Hospital Type	Hospital Type	Hospital Type	Hospital Type	Hospital Type	Hospital Type
Diagnosis Groups Proposed For Exclusion (based on ICD-10 codes)	University (main campus) (Hosp N=9) (Pt N=15,071)	AAP Level 3/4 (Hosp N=108) (Pt N=211,903)	AAP Level 2 (N=57) (Pt N=65,686)	AAP Level 1 (not Critical Access) (Hosp N=61) (Pt N=29,046)	Critical Access (Hosp N=12) (Pt N=1,684)	All Hospitals (Hosp N=238) (Pt N=308,319)
Care of Fetal anomalies	54.3	40.2	32.4	0	0	38.6

*	Hospital Type	Hospital Type	Hospital Type	Hospital Type	Hospital Type	Hospital Type
HIV	39.1	44.9	20	0	No cases	40
Severe Preeclampsia	34.8	47.8	55.6	50.6	37.5	49.6
Cardiovascular	28.9	32.1	35.6	33.3	26.7	32.7
Kidney HTN	20	25.7	75	0	No cases	30
Cerebral Thrombosis	No cases	100	0	No cases	No cases	50
Previa expanded	50	42.1	20	0	No cases	36
Low-lying placenta	61.3	54.4	47.3	57.8	100	53.5
Accreta	16.7	44	40.9	22.2	No cases	41.5
Abruption	33.3	52.4	75	100	No cases	59.3
Cord Prolapse	81.8	84.3	83.7	68.8	66.7	81.6
Vasa Previa	66.7	71.4	100	66.7	No cases	71.8
Any of the above	35.2	41.9	44.4	46.2	37.9	42.6

* Cell intentionally left empty

Note: the cesarean rate for placenta accreta may seem low but this is a term nulliparous population so most of these cases were diagnosed in the setting of retained placentas after vaginal delivery and not the very troublesome placenta accretas seen with a previa after prior cesarean birth(s). The second observation is that while there was a slightly higher rate of those complications (Table 1) at University hospitals, the cesarean rate for these complications was relatively low.

Table 2b.21.02 Cesarean Delivery Rate (%) for selected major obstetric complications (among NTSV PC-02 population). Five hospital type categories are displayed horizontally: University, AAP Level 3 / 4, AAP Level 2, AAP Level 1, Critical Access and Total. Vertically 12 Diagnosis Groups Proposed For Exclusion (based on ICD-10 codes) are displayed.

Table 2b.21.03 Revised NTSV Cesarean Delivery Rate (%) with selected major obstetric complications excluded (among NTSV PC-02 population)

*	Hospital Type	Hospital Type	Hospital Type	Hospital Type	Hospital Type	Hospital Type
Diagnosis Groups Proposed For Exclusion (based on ICD-10 codes)	University (main campus) (Hosp N=9) (Pt N=15,071)	AAP Level 3/4 (Hosp N=108) (Pt N=211,903)	AAP Level 2 (N=57) (Pt N=65,686)	AAP Level 1 (not Critical Access) (Hosp N=61) (Pt N=29,046)	Critical Access (Hosp N=12) (Pt N=1,684)	All Hospitals (Hosp N=238) (Pt N=308,319)
Baseline NTSV (PC-02)	22.7	24.6	24	25.9	21.7	24.6
Care of Fetal anomalies	22.6	24.6	23.9	25.9	21.7	24.5
HIV	22.7	24.6	24	25.9	21.7	24.6
Severe Preeclampsia	22.7	24.5	23.9	25.8	21.6	24.5
Cardiovascular	22.6	24.5	23.9	25.9	21.6	24.5
Kidney HTN	22.7	24.6	23.9	25.9	21.7	24.6

*	Hospital Type	Hospital Type	Hospital Type	Hospital Type	Hospital Type	Hospital Type
Cerebral Thrombosis	22.7	24.6	24	25.9	21.7	24.6
Previa expanded	22.7	24.6	24	25.9	21.7	24.6
Low-lying placenta	22.6	24.5	23.9	25.8	21.6	24.5
Accreta	22.7	24.6	23.9	25.9	21.7	24.6
Abruption	22.7	24.6	23.9	25.9	21.7	24.6
Cord Prolapse	22.7	24.5	23.9	25.8	21.6	24.5
Vasa Previa	22.7	24.6	24	25.9	21.7	24.6
Any of the above	22.3	24.3	23.7	25.7	21.4	24.3

* Cell intentionally left empty

Note: The exclusion of these additional complications results in a 0.3 percentage point reduction (24.6 to 24.3%) which is consistent among all hospital types. There is no evidence that any one hospital type is disadvantaged by not excluding these diagnoses. In fact, University hospitals, despite having a presumptive higher risk patient population, have lower NTSV cesarean rates both before and after the additional exclusions were considered.

Table 2b.21.03 Revised NTSV Cesarean Delivery Rate (%) with selected major obstetric complications excluded (among NTSV PC-02 population). Five hospital type categories are displayed horizontally: University, AAP Level 3 / 4, AAP Level 2, AAP Level 1, Critical Access and Total. Vertically 12 Diagnosis Groups Proposed For Exclusion (based on ICD-10 codes) are displayed.

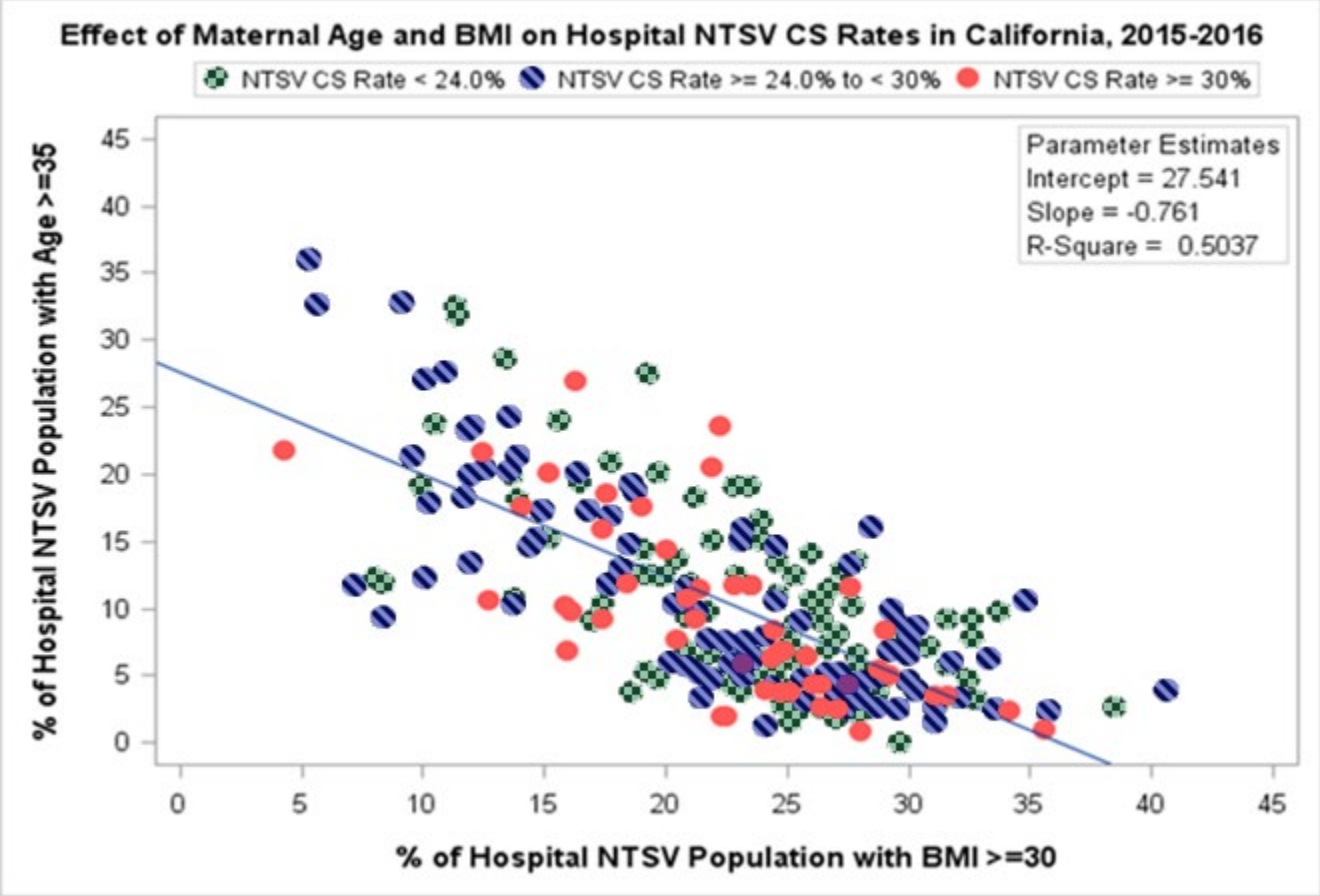
1. Armstrong, J., McDermott, P., Saade, G. R., Srinivas, S. K., Society for Maternal-Fetal Medicine Health Policy and Advocacy Committee, & Society for Maternal-Fetal Medicine Coding Committee (2017). Coding update of the SMFM definition of low risk for cesarean delivery from ICD-9-CM to ICD-10-CM. *American Journal of Obstetrics and Gynecology*, 217(1), B2–B12.e56. https://doi.org/10.1016/j.ajog.2017.04.013

CMQCC Analysis of Effects of Maternal Age and BMI on NTSV Cesarean rate

Several studies have demonstrated an effect on individual cesarean rates for both advancing maternal age and higher BMI. However, these effects on hospital NTSV cesarean rates are complex for two reasons: (1) hospitals with a birth population of high maternal age also tend to have low BMI and likewise those hospitals with low maternal age tend also to have higher BMI; (2) The actual rates for cesarean delivery in women with high maternal age or high BMI varies greatly from hospital to hospital indicating a large degree of subjectivity for the cesarean decision making, independent of the risk factor. We illustrate this is two ways, one descriptive and one analytic.

Descriptive Approach: In Figure 1 we have graphed the proportion of the hospital's birthing population that has advanced maternal age (≥35 years) versus the proportion of the hospital's population that has a pre-pregnancy BMI >30 for 242 California hospitals with an average of ≥100 annual births continually open from 2015-2016. A moderate correlation between age and BMI is noted. The hospital dots are color coded by their NTSV rate: green for <24%, blue for 24-30% and red for >30%. There are two notable observations: (1) green and red dots are widely distributed through the graph; and (2) for every Age/BMI intercept with a red dot there are multiple green dots nearby with similar Age/BMI populations. This would support the conclusion that provider/nursing practice(s) is the main driver for the variation in care noted for age/BMI and lack of need for adjustment.

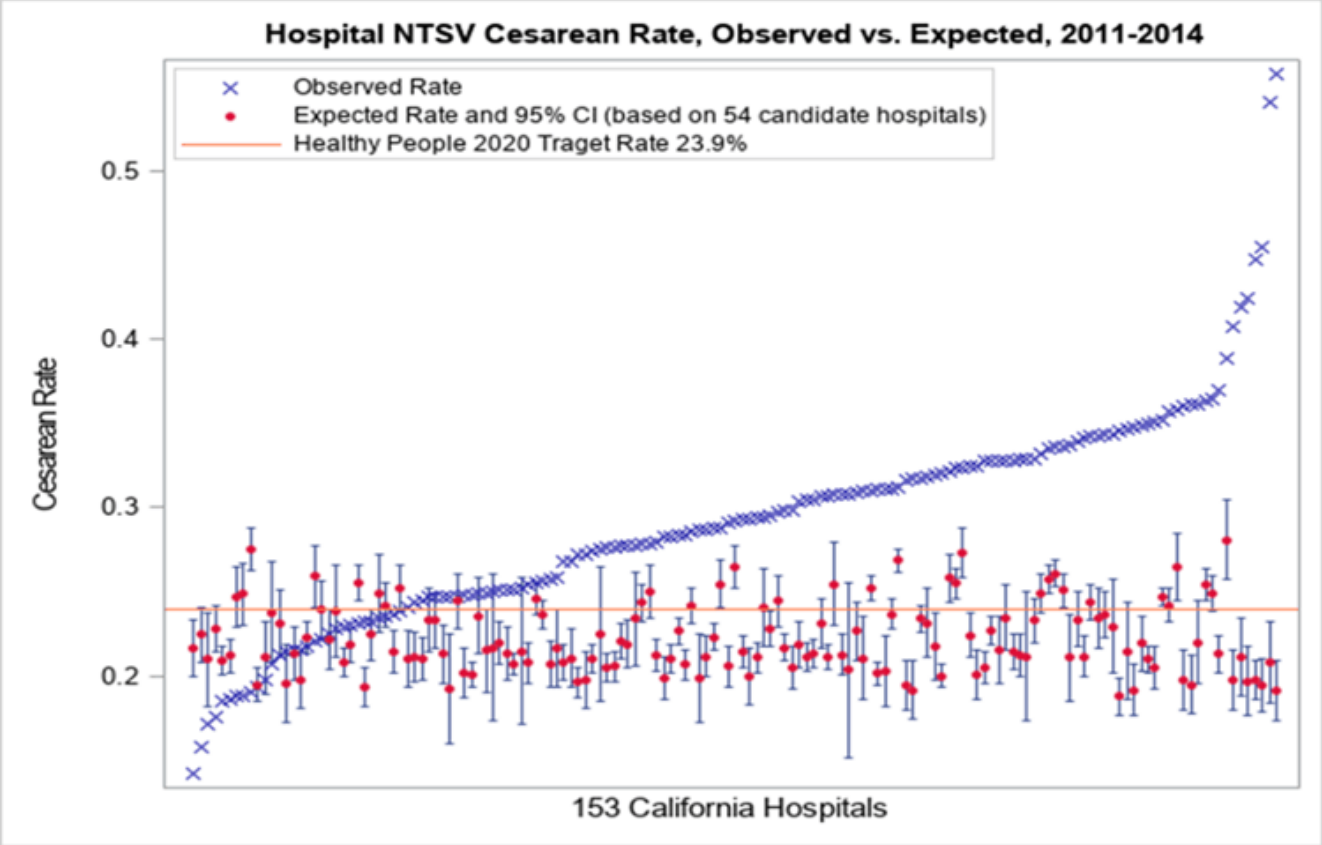
Figure 2b.21.01 Overlap of Age and BMI populations for high and low NTSV Cesarean rate hospitals



Analytic Approach: Here we ask what if we identified a set of best practice hospitals and asked what would the other hospital's NTSV cesarean rates be if they delivered in these best practice hospitals? After setting aside Kaiser facilities because of their different care model, best practice hospitals were identified by being in both the lower 50 percentile for NTSV cesarean rates and in the lower 50 percentile for unexpected newborn complications (a NQF-endorsed composite term neonatal outcome measure that is now PC-06). This population of 54 hospitals with both lower CS rates and lower rates of poor baby outcomes became the standard hospitals for the next step. We then asked what would the NTSV cesarean rate be if a given hospitals individual patients were delivered at a best practice hospital. This was achieved by propensity mapping each patient in the non-best practice facility by their age and BMI to exact matches within the best practice hospitals. Figure 2 below shows the results. The x's illustrate the variation observed among the 153 hospitals that are not the best performers (for both NTSV and unexpected newborn complications). The expected rates if those hospital's patients had been delivered at a best practice facility are shown by red dots. The results are dramatic. Nearly all of the large variation in NTSV shown by the x's has been removed and now hospitals cluster around 22% (19-25%). This indicates that physician preference and subjectivity account for most of the Age and BMI effects on NTSV cesarean rate again supporting the lack of need for adjustment for these factors.

It should be noted that this was done with a fairly generous definition of NTSV best practice-only that the hospital had to be below the mid-point which for this time period (2011-2014) was 26.1%. The current average (2018) in California is 23.4% which would give significantly lower absolute rates if repeated again. This data is under submission for publication.

Figure 2b.21.02 Observed and Expected rate of 153 California non-best performing hospitals had their patients delivered in the 54 best performing facilities.



[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]
[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]
[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]
[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]
[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]
[Response Ends]

2b.27. Provide risk model discrimination statistics.

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

[Response Begins]
[Response Ends]

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

[Response Begins]
Not applicable.
[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]
[Response Ends]

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

[Response Begins]
[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]
[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]
[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: 0471e_Feasibility Scorecard ePC02.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]
One site uses a standalone OB documentation system that does not interface completely with the electronic health record. The OB documentation comes from a 3rd party L&D system that is shared with the primary EHR system as non-discrete data. Site two's performance was directly related to this issue as the data availability of these data elements resulted in an 83% feasibility rate; however, we did not find this issue in other sites using the same primary EHR. Site 2 has identified solutions which they have implemented since the pilot testing which would allow 100% feasibility for these data elements. Additional details are provided in sections 4.08 and 4.10.
Since the measure has been implemented by 15 hospitals, we can conclude that the measure is feasible. During the reliability visits at 2 sites (7 hospitals), feasibility scorecards were completed, and the feasibility rate was found to be 98% across two electronic health record (EHR) systems (Epic and Meditech).
Feasibility Testing: We conducted a virtual EHR walkthrough session with each pilot site. The pilot site shared their screen while navigating through their EHR system as the measure data elements, specifications, and clinical workflows were discussed. Using the NQF's eCQM Feasibility Scorecard template, a scorecard was completed for each pilot site during this time. The feasibility scorecard results were analyzed for each site and aggregated across all pilot sites (see Table 3.06.01

below). Each data element score was examined within each of the domains (see Table 3.06.02 below). Highly feasible was defined as receiving the maximum score of 1 within the domains and was expressed as a percentage.

Table 3.06.01 Overall Feasibility Rates

PILOT SITE	FEASIBLITY RATE
1	100%
2	95%
Overall	98%

Table 3.06.01 Overall Feasibility Rates show the feasibility rate for Pilot Site 1 (100%), Pilot Site 2 (95%) and Overall (98%).

Table 3.06.02 Feasibility Rates by Domain

PILOT SITE	DATA AVAILABILITY	DATA ACCURACY	DATA STANDARDS	WORKFLOW
1	100%	100%	100%	100%
2	83%	97%	100%	100%
Overall	92%	98%	100%	100%

Table 3.06.02 Feasibility Rates by Domain show the feasibility rate for Pilot Site 1, Pilot Site 2, and Overall broken down by the 4 domains of Data Availability, Data Accuracy, Data Standards, and Workflow. Pilot Site 1 scored 100% on all 4 domains. Pilot Site 2 scored 83% on Data Availability, 97% on Data Accuracy and 100% on Data Standards and Workflow.

[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]
There are no fees or licensing requirements to use The Joint Commission performance measures, all of which are in the public domain.
[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 and 81% of beds. Greater than 2500 accredited US hospitals nationwide have maternity services. The Cesarean Birth measure is available to these hospitals to meet ORYX requirements.
- **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. Data collected from the accredited hospitals are analyzed for tends and benchmarks.
- **Geographic area and number and percentage of accountable entities and patients included:** The Joint Commission accredits 63% of hospitals and 81% of beds. Greater than 2500 accredited US hospitals nationwide have maternity services. The Cesarean Birth measure is available to these hospitals to meet ORYX requirements.
- **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. Data collected from the accredited hospitals are analyzed for tends and benchmarks. Organizations are provided this data for use in internal quality improvement.
- **Geographic area and number and percentage of accountable entities and patients included:** The Joint Commission accredits 63% of hospitals and 81% of beds. Greater than 2500 accredited US hospitals nationwide have maternity services. The Cesarean Birth measure is available to these hospitals to meet ORYX requirements.
- **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]
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]
Yearly educational webinars are provided through The Joint Commission's Pioneers in Quality program. These webinars provide measure specification review, updates and offer a Q&A opportunity for audience members.
The Joint Commission developed dashboards as part of an ongoing project to provide continuous customer engagement. 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]
The Joint Commission utilizes an automated feedback system with access 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.
In 2020, The Joint Commission introduced the Cesarean Birth measure (ePC02) as one of the available eCQMs hospitals could choose for data submission to meet ORYX requirements. 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 6 sites consisting of 15 hospitals submitted production data for one quarter of calendar year 2020. These data (which will be referred to as production data) were used for all of the testing provided with the exception of validity testing, which used a subset of the six sites. TJC reached out to all 15 hospitals to recruit sites willing to participate in validity testing on the data submitted. Two pilot sites (7 hospitals) volunteered. One site is a system representing 6 hospitals where the Epic system is used. The 7th hospital is a stand-alone facility that uses Meditech. The two pilot sites (7 hospitals) provided feedback during feasibility (NQF scorecard) and validity testing. This data will be referred to as pilot test data. The virtual pilot testing sessions were used to 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. Email correspondence and live Q&A during the virtual testing sessions were part of the feedback process. See 4a.08 for feedback details.

[Response Ends]

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

[Response Begins]
Feasibility Test Feedback: Feedback obtained during feasibility testing indicated that only one data element (Assessment, Performed: Estimated Gestational Age at Delivery, author Datetime) was problematic in that the results were present but

the estimated gestational age author data/time was not consistently captured accurately. The other data elements which led to possible feasibility issues were as follows:
Assessment, Performed: Time of delivery, authorDatetime
Assessment, Performed: Births.preterm, authorDatetime
Assessment, Performed: Births.preterm, result
Assessment, Performed: Births.term, authorDatetime
Assessment, Performed: Births.term, result
These data elements provided accurate data, but they were from OB documentation which comes from a 3rd party L&D system that is shared with the primary EHR system as non-discrete data.
Validity Test Feedback: Feedback obtained during validation of production data showed an overall data element agreement rate of 92%. At pilot site one, six cases were miscoded as malpresentation of the fetus when the fetus was actually in the vertex position. At the second site, the hospital uses a 3rd party L&D system for OB documentation that does not interface completely with the Meditech Electronic Health Record. A .pdf report is available in Meditech in non-discrete fields. Most mismatches were in the Delivery Date/Time, Estimated Gestational Age, Gravida, Para, Preterm or Term Birth fields.

[Response Ends]

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

[Response Begins]
Feedback has not been obtained by other users. The Joint Commission's online Performance Measurement Network Q&A Forum remains available to users to provide feedback.
[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]
Here is a summary of feedback received during testing of the production data and how the feedback was interpreted and used to improve the measure.
The overall lower Kappa levels for Site 2 were due to 10 specific data elements. Issues with 3 of those data elements we feel are resolved as the current version of ePC02 has updated author date/time to relevant date time and allows 2 ways to determine gestational age (calculated using date of delivery and estimated due date OR reported EGA). These data elements are shared with ePC-07 and ePC07 pilot testing showed excellent match rates in 3 EHRs (EPIC, Cerner, Meditech) with 94-98 percent match rates.

- Assessment, Performed: Date and time of obstetric delivery, Author Date Time has been replaced with Assessment, Performed: Date and time of obstetric delivery, relevant date/time which had a 98% match rate.
- Assessment, Performed: Estimated Gestational Age at Delivery, Author Date Time rate has been replaced with Assessment, Performed: Estimated Gestational Age at Delivery, relevant date/time which had a 94% match rate.
- Assessment, Performed: Estimated Gestational Age at Delivery, result had a 96% match rate.

The other 7 data elements are related to Preterm, Term, Parity, results and their associated author date time and Gravidity author date time only (Gravida result had a rate of 91.2 at site 2 and 100% at site1).

- Assessment, Performed: Births.preterm - Author Date Time
- Assessment, Performed: Births.preterm - Result
- Assessment, Performed: Births.term - Author Date Time
- Assessment, Performed: Births.term - Result
- Assessment, Performed: Parity - Author Date Time
- Assess Perf Parity - Result
- Assessment, Performed: pregnancies (gravida) - Author Date Time

57% of the mismatch for these data elements were due to missing data because Site 2 used a 3rd party L&D system for OB documentation which did not interface completely with the electronic health record (Meditech). The OB documentation was present in Meditech in non-discrete fields in a .pdf format. The site has since implemented changes where the data is now stored in discrete fields and therefore the data is able to be captured by the eCQM, however, the site was unable to submit updated data in time for NQF submission.

We feel confident that these data elements are able to be accurately abstracted in an EHR system as evidenced by site 1's 96-100% match rates on all 10 of these data elements. The ePC02 measure logic only requires parity OR gravidity OR preterm and term not all 4 data elements together. When accounting for the root cause site 2 had low Kappas which were able to be mitigated in the future and using other available evidence which shows high kappas in multiple EHRs for the shared data elements, we feel that overall, this measure is valid and able to capture differences in performance. As described in 4a.08, at pilot site one, six cases were miscoded as malpresentation of the fetus when the fetus was actually in the vertex position. No update is needed to the measure specifications. The hospital planned to work with their coding staff to rectify this situation.

[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]
Yearly trends are not available as TJC has only received one year's worth of data at this point (2020 discharges).
The measure will assist health care organizations to track nulliparous patients with live term singleton newborns in vertex position delivering by cesarean birth to reduce the occurrence. A reduction in the number of nulliparous patients with live term singleton newborns in vertex position (NTSV) delivering by cesarean birth will result in increased patient safety, a substantial decrease in maternal and neonatal morbidity and substantial savings in health care costs. See 1b.01 for additional details.

[Response Ends]

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

[Response Begins]
As described in 4a.08 One hospital uses a stand-alone OB documentation system that does not interface completely with the Meditech Electronic Health Record. A .pdf report is available in Meditech containing non-discrete fields. The hospital has developed a mitigation plan to ensure that the necessary data elements are available in discrete fields. No other unexpected findings or unintended impacts were identified.
[Response Ends]

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

[Response Begins]
As described in 4a.08, at a second site, six cases were miscoded as malpresentation of the fetus when the fetus was actually in the vertex position. The hospital planned to work with their coding staff to rectify this situation which will result in improved coding practices
[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]
0471: PC-02 Cesarean Birth
[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]
Yes
[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]
N/A. Measure is harmonized.
[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]
There are no other NQF endorsed competing measures or non-endorsed facility level measures for cesarean birth. 0471e Cesarean Birth is harmonized with the NQF endorsed chart-based measure 0471 Cesarean Birth. We developed the

eCQM version of PC02 to reduce administrative burden for sites able to report it and to encourage the use of eCQMs. We accept either eCQM or chart-abstracted data (or both) for The Joint Commission accreditation program. Thirteen Joint Commission accredited hospitals submitted PC-02 data for both the eCQM and chart-abstracted measures in calendar year 2020. The ePC-02 rates for the 13 hospitals who submitted both eCQM and chart-abstracted measure results to The Joint Commission for 2020 discharges were correlated at 0.88 which is strong and is statistically significant ($p<0.01$). The eCQM data for this correlation came from 2 EHR systems EPIC and Meditech. This eCQM is important to allow hospitals capable of submitting the electron version of the Cesarean Birth measure to do so which can decrease burden of manual abstraction, increase efficiencies, and improve experience.

[Response Ends]