NATIONAL QUALITY FORUM

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 0469e

Corresponding Measures: 0469

De.2. Measure Title: PC-01 Elective Delivery

Co.1.1. Measure Steward: The Joint Commission

De.3. Brief Description of Measure: This measure assesses patients with elective vaginal deliveries or elective cesarean births at >= 37 and < 39 weeks of gestation completed. This measure is part of a set of four nationally implemented measures that address perinatal care (PC-01: Elective Delivery, ePC-01: Elective Delivery; PC-02: Cesarean Birth, ePC-02: Cesarean Birth will be added as an eCQM 1/1/2020; PC-05: Exclusive Breast Milk Feeding, ePC-05: Exclusive Breast Milk Feeding; PC-06 Unexpected Complications in Term Newborns was added 1/1/2019).

PC-01: Elective Delivery is one of three measures in this set that have been re-engineered as eCQMs (ePC-01 elective Delivery, ePC-05 Exclusive Breast Milk Feeding and ePC-02 Cesarean Birth).

A reduction in the number of non-medically indicated elective deliveries at >=37 to <39 weeks gestation will result in a substantial decrease in neonatal morbidity and mortality, as well as a significant savings in health care costs. In addition, the rate of cesarean deliveries should decrease with fewer elective inductions resulting in decreased length of stay and health care costs (AAFP, 2000). The measure will assist health care organizations (HCOs) to track non-medically indicated early term elective deliveries and reduce the occurrence.

American Academy of Family Physicians. (2000). Tips from Other Journals: Elective induction doubles cesarean delivery rate, 61, 4.Retrieved December 29, 2008 at: http://www.aafp.org/afp/20000215/tips/39.html.

1b.1. Developer Rationale: For almost 3 decades, the American College of Obstetricians and Gynecologists (ACOG) and the American Academy of Pediatrics (AAP established and followed the standard requiring 39 completed weeks gestation prior to ELECTIVE delivery, either vaginal or operative (ACOG, 1996). A survey conducted in 2007 of almost 20,000 births, in HCA hospitals throughout the U.S. carried out in conjunction with the March of Dimes at the request of ACOG, revealed that almost one-third of all babies delivered in the United States are electively delivered with 5% of all deliveries in the U.S. delivered in a manner violating ACOG/AAP guidelines. Most of these are for convenience and result in significant short-term neonatal morbidity, such as neonatal intensive care unit admission rates of 13- 21% (Clark et al., 2009). A reduction in the number of non-medically indicated elective deliveries at >=37 to <39 weeks gestation will result in a substantial decrease in neonatal morbidity and mortality, as well as a significant savings in health care costs. In addition, the rate of cesarean deliveries should decrease with fewer elective inductions resulting in decreased length of stay and health care costs.

The measure will assist health care organizations (HCOs) to track non-medically indicated early term elective deliveries and reduce the occurrence.

Sources

- According to Glantz (2005), compared to spontaneous labor, elective inductions result in more cesarean births and longer maternal length of stay. The American Academy of Family Physicians (2000) also notes that elective induction doubles the cesarean delivery rate. Repeat elective cesarean births before 39 weeks gestation also result in higher rates of adverse respiratory outcomes, mechanical ventilation, sepsis and hypoglycemia for the newborns (Tita et al., 2009).
- American College of Obstetricians and Gynecologists. (November 1996). ACOG Educational Bulletin.
- American Academy of Family Physicians. (2000). Tips from Other Journals: Elective induction doubles cesarean delivery rate, 61, 4.Retrieved December 29, 2008 at: http://www.aafp.org/afp/20000215/tips/39.htm
- Clark, S., Miller, D., Belfort, M., Dildy, G., Frye, D., & Meyers, J. (2009). Neonatal and maternal outcomes associated with elective delivery. [Electronic Version]. Am J Obstet Gynecol. 200:156.e1-156.e4.
- Glantz, J. (Apr.2005). Elective induction vs. spontaneous labor associations and outcomes. [Electronic Version]. J Reprod Med. 50(4):235-40.
- Tita, A., Landon, M., Spong, C., Lai, Y., Leveno, K., Varner, M, et al. (2009). Timing of elective repeat cesarean delivery at term and neonatal outcomes. [Electronic Version]. NEJM. 360:2, 111-120.

S.4. Numerator Statement: Inpatient hospitalizations for patients with elective deliveries by either:

- Medical induction of labor while not in labor prior to the procedure

- Cesarean birth while not in labor and with no history of a prior uterine surgery

S.6. Denominator Statement: Inpatient hospitalizations for patients delivering newborns with >= 37 and < 39 weeks of gestation completed.

S.8. Denominator Exclusions: Inpatient hospitalizations for patients with conditions possibly justifying elective delivery prior to 39 weeks gestation.

De.1. Measure Type: Process

S.17. Data Source: Electronic Health Data, Electronic Health Records, Other

S.20. Level of Analysis: Facility, Other

IF Endorsement Maintenance – Original Endorsement Date: Oct 25, 2016 **Most Recent Endorsement Date:** Oct 25, 2016

IF this measure is included in a composite, NQF Composite#/title:

IF this measure is paired/grouped, NQF#/title:

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? Not Applicable

Preliminary Analysis: Maintenance of Endorsement

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria (“maintenance”). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence

Maintenance measures – less emphasis on evidence unless there is new information or change in evidence Since the prior evaluation.

1a. Evidence. The evidence requirements for a *structure, process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured. For measures derived from patient report, evidence also should demonstrate that the target population values the measured process or structure and finds it meaningful.

The developer provided the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☐ **Yes** ☒ **No**

Summary of prior review in 2016

- This process measure was last evaluated in 2016 based primarily on the following evidence: American College of Obstetricians and Gynecologists (ACOG) Practice Bulletin No. 107—systematic review of literature (evidence not graded, recommendation Level II).
- Elective deliveries performed at =>39 weeks gestation result in improved maternal and neonatal outcomes and will result in a substantial decrease in cesarean sections and neonatal morbidity, as well as substantial savings in healthcare costs.
- The body of evidence reviewed consistently supported the benefit of reduction of non-medically indicated early term elective deliveries.
    - All studies demonstrated an increase in the number of neonatal morbidities associated with early term deliveries, subsequent reduction of elective non-medically indicated deliveries reduces harm to the neonate
    - All studies demonstrated similar findings related to the direction of effect, though the magnitude varied from study to study.
- The developer reported Quantity = High; Quality = Moderate; Consistency = High

Changes to evidence from last review

☐ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**

☒ **The developer provided updated evidence for this measure:**

Updates:

- 2019 ACOG Committee Opinion—Avoidance of nonmedically indicated early-term deliveries and associated neonatal morbidities.
- Five recommendations by ACOG and Society of Fetal Medicine made, including: Nonmedically indicated delivery, including cesarean delivery, inductions of labor, and cervical ripening should not occur before 39 0/7 weeks of gestation (evidence not graded, no recommendation level/grade).
- Systematic review of literature (24 studies) found late-preterm and early-term children have lower performance scores across (a range of cognitive and educational measures compared with their full-term peers. Non-respiratory morbidities also are increased in early-term deliveries; documentation of fetal pulmonary maturity does not justify an early nonmedically indicated delivery.

Exception to evidence

Not applicable

Question for the Committee:

- *The developer provided updated evidence for this measure that is directionally consistent and strengthens the evidence for the previous NQF review. Does the Committee agree there is no need for repeat discussion and vote on Evidence?*

Guidance from the Evidence Algorithm

Process measure based on systematic review (Box 3) → Evidence not graded (Box 7) → Developer assesses Quantity: high; Quality: moderate; Consistency: high (Box 9) → Moderate.

The highest possible rating is Moderate.

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

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

Maintenance measures – increased emphasis on gap and variation

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

No data were available in 2016, as it was a new e-measure. For 2018 data (N=131):

- Mean 17.6%, SD 21.6%
- Distribution
  - 10th Percentile= 0%
  - 25th Percentile= 0%
  - 50th Percentile= 13.3%
  - 75th Percentile= 25.0%
  - 90th Percentile= 41.4%

Trend data were not provided.

Disparities

Based on 2018 discharges:

- Performance by age category

Age	Rate (%)
<20	14.6
20-24	18.1
25-29	18.7
30-34	22.5
35-39	18.8
40+	32.7

- Performance by Hispanic Ethnicity

Hispanic Ethnicity	Rate (%)
No	23.3
Yes	20.8

- Performance by race

Race	Rate (%)
White	22.3
African American	18.3
American Indian	41.7
Asian	16.2

Pacific Islander	25.0
Other	15.0

Question for the Committee:

- *Is there a gap in care and/or disparities that warrant a national performance measure? Of note, the performance rates on the "paper" version of the measure differ significantly (mean 1.7%).*

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

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

1a. Importance to Measure and Report

Comments:

** See comments for 0469

** not aware of new uncited evidence

** The evidence has not changed substantially and I agree that there is no need for repeat discussion and vote on Evidence.

1b. Performance Gap

Comments:

** More significant gap in care with this measure as compared to 0469.

** Yes- curious why this data is so different from the related, non-e measure

** There are still opportunities for improvement on this measure.

1b. Disparities

Comments:

** More disparities with this measure than with 0469

** yes, by age and race. Startling rate for American Indian.

** I would like to see an analysis of significance of variation based on age, ethnicity, and race.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

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

Reliability

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

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

Validity

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

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

eCQM Technical Advisor (s) review:

Submitted measure is an HQMF compliant eCQM	The submitted eCQM specifications follow the industry accepted format for eCQM (HL7 Health Quality Measures Format (HQMF)). **HQMF specifications** ☒ **Yes** ☐ **No**
Documentation of HQMF, QDM, or CQL limitations	N/A – All components in the measure logic of the submitted eCQM are represented using the HQMF, QDM, or CQL standards
Value Sets	The submitted eCQM specifications uses existing value sets when possible and uses new value sets that have been vetted through the VSAC
Measure logic is unambiguous	Submission includes test results from a simulated data set demonstrating the measure logic can be interpreted precisely and unambiguously. – this includes 100% coverage of measured patient population • Submission includes simulated data set results demonstrating unit testing covering 100% of the measure logic. Testing with pass/fail test cases for each population.
Feasibility Testing	• Feasibility assessment indicated that 15 data elements could not be assessed for accuracy. (See item #20 below.) Additionally, two data elements were missing from the feasibility assessment. For each of the elements, the developer mentioned that while it could not assess the accuracy, it believes the element to be accurate due to the measure logic being derived from the chart abstracted version of this measure.

Questions for the Committee regarding reliability:

- *Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?*
- *The developer notes several changes to the specifications to improve alignment with the "paper" measure. Does the Committee wish to discuss with the developer the degree to which non-alignment affects the measure scores for each version?*
- *The developer uses data element validity to stand in for empirical reliability testing, as permitted by NQF. In doing so, it found near perfect agreement for three of six data elements, but moderate to poor reliability for the other three (medical induction of labor, active labor, and prior uterine surgery); the developer offers an explanation of these results, including that the three are less critical. The Committee should discuss whether these results and the justification are sufficient for a MODERATE or LOW rating.*

Questions for the Committee regarding validity:

- *The eCQM feasibility assessment indicated that certain data elements could not be assessed for accuracy. For each of the elements, the developer mentioned that while it could not assess the accuracy, it believes the element to be accurate due to the measure logic being derived from the chart abstracted version of this measure. How do these data elements impact the validity of the measure?*

- *Two data elements are missing from the feasibility assessment. Does the Committee wish to discuss with the developer the impact, if any, on validity?*
- *Does the Committee wish to discuss how the data elements are feasible in the context of the measure logic?*
- *Does the Committee wish to discuss with the developer how the data elements are used in computation of the measure?*
- *Does the Committee have concerns about missing data?*
- *Are you satisfied with the score-level validity analyses for the measure?*

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

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

RATIONALE:

- The data element validity testing, which stands in for data element reliability testing, found near perfect agreement for three of six critical data elements, but moderate to poor reliability for the other three; the developer offers an explanation of these results, including the assessment that the three "poor" kappa data elements (medical induction of labor, active labor, and prior uterine surgery) are less critical. The Committee should discuss whether the three moderate/poor data elements are less critical and so a rating of MODERATE or LOW should be assigned.

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 0496e

Measure Title: PC-01 Elective Delivery e

Type of measure:

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

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

Data Source:

☐ **Claims** ☐ **Electronic Health Data** ☒ **Electronic Health Records** ☐ **Management Data**
☐ **Assessment Data** ☐ **Paper Medical Records** ☐ **Instrument-Based Data** ☐ **Registry Data**
☐ **Enrollment Data** ☐ **Other**

Level of Analysis:

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

Measure is:

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

RELIABILITY: SPECIFICATIONS

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

   Submission document: "MIF_0469e" document, items S.1-S.22

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

   - Submitted measure specification follows eCQM industry specs

- Submitted measure specification is fully represented and is not hindered by any limitations in the eCQM industry specs
- Specifications have been updated since last NQF review to improve alignment with the chart-abstracted measure:
  - In 2017, the value set, Conditions Possibly Justifying Elective Delivery, was updated to include concepts representing Stillbirth and History of Stillbirth to better align with the chart abstracted measure. Due to the QDM 4.3 changes, the datatype was replaced for 'Physical Exam, Performed' to 'Assessment, Performed' for the data elements of Gestational Age, Time of Delivery, and Labor. Additionally, the timing logic for 'Time of Delivery' was updated to: < 1 day(s) starts before or concurrent with start of (´Occurrence A of Assessment, Performed: Time of Delivery´ starts during Occurrence A of EncounterInpatient) to better align with clinical intent.
  - In 2018, two value sets for procedures performed were added to the history of prior uterine surgery procedure to better align with the clinical intent and the chart-abstracted version of the measure. Measure specifications also were converted from the Quality Data Model (QDM) measure logic to the Clinical Quality Language (CQL).
  - In 2019, results from an empirical analysis comparing the disparities between the chart abstracted and eCQM data created the need to update the timing statement related to the documentation of the Estimated Gestational Age (EGA) to capture the last EGA assessment within one day or less prior to or at the same time of delivery.

2. **Briefly summarize any concerns about the measure specifications.**
   - As noted in Item #20 (below), 15 items were not assessed for accuracy and two data elements were not included in the feasibility assessment.

RELIABILITY: TESTING

Submission document: "MIF_0469e" document for specifications, testing attachment questions 1.1-1.4 and section 2a2

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

   Submission document: Testing attachment, section 2a2.2
   - Empirical validity testing of patient-level data was conducted and, per NQF policy, may be used to fulfill reliability testing.
7. **Assess the results of reliability testing**

   Submission document: Testing attachment, section 2a2.3
   - See summary under Validity Testing @ Item 21.
8. **Was the method described and appropriate for assessing the proportion of variability due to real differences among measured entities?** NOTE: If multiple methods used, at least one must be appropriate.

   Submission document: Testing attachment, section 2a2.2

   ☐ **Yes**

   ☐ **No**

   ☒ **Not applicable** (score-level testing was not performed)
9. **Was the method described and appropriate for assessing the reliability of ALL critical data elements?**

Submission document: Testing attachment, section 2a2.2

☒ **Yes**

☐ **No**

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

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

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

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

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

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

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

    - Validity testing at the data element level of patient-level was performed. Per NQF guidelines, this approach may be used to demonstrate reliability. The highest possible rating in the absence of score-level reliability testing is MODERATE.
    - The data element validity testing found near perfect agreement for three of six critical data elements, but moderate to poor reliability for the other three; the developer offers an explanation of these results, including the assessment that the three "poor" data elements are less critical. The Committee should discuss whether these results and the justification are sufficient for a different rating. See question 20 for additional details.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

    Submission document: Testing attachment, section 2b2.

    - In the current submission, two exclusions (at least one condition possibly justifying elective delivery; gestational age <37 or >=39) were empirically tested for impact. "History of Stillbirth," tested in the "paper measure," was not tested for this eMeasure. The developer provided a rationale for each exclusion and the percentage lost to the exclusions, which are not mutually exclusive. The developer tested other exclusions in the 2016 submission.
    - Is the lack of empirical testing for "History of Stillbirth" in this eMeasure as compared to the paper measure a concern?

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

    Submission document: Testing attachment, section 2b4

    - To demonstrate meaningful differences in performance, the developer calculated a funnel plot for the annual hospital rates of the measure, where the observed measure is plotted against a measure of its precision, so that the control limits form a 'funnel' around the target outcome. It superimposes the 95 per cent (≈two standard deviation) and 99.8 per cent (≈three standard deviation) prediction limits over this plot around the overall measure rate; those rates lying outside the confidence limits are identified as outliers. (Spiegelhalter, DJ. Funnel plots for comparing institutional performance. *Statistics in Medicine*. 2005; 24:1185–1202.)
    - The developer reported that of 131 hospitals, 14 hospitals were identified as outliers, with rates beyond the two standard deviation upper limit. Eight hospitals were high outliers, with rates beyond three standard deviations.

- The developer stated that the results indicate significant differences in performance among hospitals and an appreciable number of hospitals that are not within the expected level of variability and differ significantly from the mean overall rate.
- No concerns.

14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**
Submission document: Testing attachment, section 2b5

- Not applicable.

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

Submission document: Testing attachment, section 2b6.

- The developer used the data from the data element validity testing to identify data elements that were missing in the eCQM, but present in the chart (14 hospitals, 14,181 patients). For each data element, a mixed model logistic regression was fit to the data, with the dependent variable being whether the data element was missing or not. A chi-squared independence test was calculated to determine if there was significant between hospital variability in the missing data rates.
- The developer examined six elements: Gestational age; conditions justifying elective delivery; medical induction of labor; cesarean birth; active labor; and prior uterine surgery.
  - The developer found significant differences in missing data rates across hospitals for all the data elements. Hospitals with 100% gestational age missing data rates would not have any cases falling in the measure and would therefore not have a measure rate to report.
  - The developer noted that its analysis was based on 2018 data, and that changes have been made to the measure in 2019 to improve the data capture rate of gestational age.
    - Based on an empirical analysis comparing the disparities between the chart abstracted and eCQM data., the timing statement related to the documentation of the Estimated Gestational Age (EGA) was updated to capture the last EGA assessment within one day or less prior to or at the same time of delivery.
    - For prior uterine surgery and active labor, missing data would result in cases being misidentified in the numerator with rates that are too high.
    - The developer posits that it expects the missing rate for these data elements to decrease with time as hospitals gain more experience with reporting this measure. It examined data from 2017 and matched it to the same hospitals in 2018 and noted that, on average, the rates are improving.

16. **Risk Adjustment**

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

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

☐ Yes ☐ No ☒ Not applicable

16c. **Social risk adjustment:**

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

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

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

16d.**Risk adjustment summary:**

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

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

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

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

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

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

- Process measure, not risk adjusted

VALIDITY: TESTING

17. **Validity testing level: ☐ Measure score ☐ Data element ☒ Both**
18. **Method of establishing validity of the measure score:**

☐ Face validity

☒ Empirical validity testing of the measure score

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

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

Submission document: Testing attachment, section 2b2.2

- Validity testing of critical data elements was performed by comparing eCQM data to the corresponding chart-based data that was submitted on the same patient.
  - Patient-level data were matched to chart-based data (matched using hospital ID, admission date, discharge date and birth date), and each data element was compared between the eMeasure data and the corresponding chart data. The developer noted that previous submissions had demonstrated that the chart data had a high degree of data element reliability.
    - Sensitivity, specificity and kappa statistics were used to measure the agreement between the two data sources, with the chart data considered to be the golden standard.
    - The data elements birth date, admission date and discharge date used in the matching were assumed to be correct and were therefore used to measure agreement.
  - For the score-level empirical testing, the developer used construct validity, hypothesizing that this measure would correlate positively with the paper measure. This measure also was correlated with the developer's other measures of perinatal care.
    - Since a lower rate for this measure = higher quality, the developer hypothesized the measure would correlate negatively to other perinatal care measures where a high rate is desirable (chart/paper measure Exclusive Breast Feeding and PC-05 and paper version of this measure).
    - The developer stated that a correlation of 0.1 - 0.3 was considered weak, 0.3 - 0.5 was considered moderate, and over 0.5 was considered strong. A kappa score above 0.6 was considered good and above 0.8 was considered excellent.

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

Submission document: Testing attachment, section 2b2.3

For data element validity testing:

- The developer noted that three of the critical data elements (gestational age in range; conditions possibly justifying elective delivery; and cesarean birth) show almost perfect agreement, with kappa scores above 0.90. The developer states that the other data elements (medical induction of labor, active labor, and prior uterine surgery) are used later in the algorithm and impact relatively fewer cases. The developer notes that kappas for these data elements have fair to poor agreement because the eCQM cannot identify the presence of the data element. The developer hypothesizes that this is perhaps due to workflow and data capture issues in the eCQM.
- The developer stated it believes the most critical data elements required for the measure show validity. It notes, however, that the fair to poor agreement on the other data elements has a "large" impact on the identification of numerator cases. (NQF staff note that mean performance in 2018 on the chart measure was 1.7%, SD 2.8%; mean performance on the eCQM measure was 17.6%, SD

21.6%.) The developer states that it provides an annual educational webinar to improve data capture for the electronic perinatal care measures.

- For its eCQM testing, 15 data elements were not assessed for accuracy, as follows:
  - 19. "Procedure, Performed: Delivery Procedures" using "Delivery Procedures Grouping Value Set (2.16.840.1.113762.1.4.1045.59
  - 15. "Physical Exam, Performed: Time of Delivery" using "Time of Delivery SNOMEDCT Value Set (2.16.840.1.113762.1.4.1045.28
  - 13. "Physical Exam, Performed: Estimated Gestational Age at Delivery" using "Estimated Gestational Age at Delivery SNOMEDCT Value Set (2.16.840.1.113762.1.4.1045.26)"
  - 4. "Diagnosis: Conditions Possibly Justifying Elective Delivery Prior to 39 Weeks Gestation" using "Conditions Possibly Justifying Elective Delivery Prior to 39 Weeks Gestation (2.16.840.1.113883.3.117.1.7.1.286)"
  - 9. "Diagnosis: Uterine Window" using "Uterine Window (2.16.840.1.113883.3.117.1.7.1.137)"
  - 8. "Diagnosis: Uterine Rupture" using "Uterine Rupture (2.16.840.1.113762.1.4.1110.16)"
  - 7. "Diagnosis: Perforation of Uterus" using "Perforation of Uterus (2.16.840.1.113762.1.4.1110.14)"
  - 6. "Diagnosis, Inactive: Cornual Ectopic Pregnancy" using "Cornual Ectopic Pregnancy SNOMEDCT Value Set (2.16.840.1.113762.1.4.1045.27)"
  - 22. "Procedure, Performed: Transabdominal Cerclage" using "Transabdominal Cerclage SNOMEDCT Value Set (2.16.840.1.113762.1.4.1110.2)"
  - 21. "Procedure, Performed: Myomectomy" using "Myomectomy SNOMEDCT Value Set (2.16.840.1.113883.3.117.1.7.1.422)" 20. "Procedure, Performed: Medical Induction of Labor" using "Medical Induction of Labor Grouping Value Set (2.16.840.1.113883.3.117.1.7.1.288)"
  - 18. "Procedure, Performed: Classical Cesarean Birth" using "Classical Cesarean Birth Grouping Value Set (2.16.840.1.113883.3.117.1.7.1.421)"
  - 17. "Procedure, Performed: Cesarean Birth" using "Cesarean Birth Grouping Value Set (2.16.840.1.113883.3.117.1.7.1.282)"
  - 16. "Procedure, Performed: Artificial Rupture of Membranes" using "Artificial Rupture of Membranes SNOMEDCT Value Set (2.16.840.1.113762.1.4.1045.57)"
  - 14. "Physical Exam, Performed: Labor" using "Labor SNOMEDCT Value Set (2.16.840.1.113883.3.117.1.7.1.281)" 12. "Medication, Administered: Oxytocin" using "Oxytocin RXNORM Value Set (2.16.840.1.113762.1.4.1045.55)"
  - 11. "Medication, Administered: Dinoprostone" using "Dinoprostone RXNORM Value Set (2.16.840.1.113762.1.4.1045.56)"
- For its eCQM testing, two data elements were missing from the feasibility assessment, as follows:
  - "Procedure, Performed: Metroplasty" using "Metroplasty (2.16.840.1.113762.1.4.1110.25)"
  - "Procedure, Performed: Uterine Horn" using "Uterine Horn (2.16.840.1.113762.1.4.1110.24)"

For score level validity testing:

- The developer correlated this measure with other measures of perinatal quality. It reported that except for the correlation between this measure and PC-05 Exclusive Breast Milk Feeding were in the expected direction. The developer further noted that although it hypothesized that this measure and PC-05 would be in the negative direction, it was not significantly greater than zero (correlation coefficient of 0.023). The developer further notes that the perinatal care measures assess different components of perinatal care and would not be expected to be more than weakly correlated since perinatal care is multidimensional.

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

    Submission document: Testing attachment, section 2b1.

    ☒ **Yes**

    ☐ **No**

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

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

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

☐ **Yes**

☒ **No**

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

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

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

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

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

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

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

- Construct validity against a more global measure (e.g., CMS's Five-Star program or HCAHPS) would be a stronger approach.
- The problem of missing data does not appear to be fully resolved.
- The moderate/poor kappas for three of six data elements for the data element validity testing are raised as part of the reliability rating. The MODERATE validity rating is based on the score-level testing.

ADDITIONAL RECOMMENDATIONS

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

- Two data elements are missing from the eCQM feasibility assessment. Does the Committee wish to discuss with the developer the impact, if any, on validity? Does the Committee wish to discuss with the developer how the data elements are feasible in the context of the measure logic?
- Three data elements are in the eCQM assessment, but not in the measure. Does the Committee wish to discuss this with the developer? Does the Committee wish to discuss with the developer how the data elements are used in computation of the measure?
- Does the Committee have concerns about missing data?

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

2a1. Reliability – Specifications
Comments:

** several exclusion elements may be difficult to obtain. Reliability is impacted by accurate coding of exclusion criteria.

** no concerns

** I question reliability in coding of some of the Conditions Possibly Justifying Elective Delivery (e.g., logistic

and psychosocial indicators), as they are not well defined and there is not consistent education regarding coding these elements.

2a2. Reliability – Testing

Comments:

** See #6.

** the less reliable data elements still appear to be important, so I am concerned about the implementation of these data elements

**The developer feels the reliability of this measure will improve with time and education.

2b1. Validity –Testing

Comments:

** Validity may be challenging if documentation and coding improvements are needed, regardless if appropriate care was delivered.

** empiric testing for "history of stillbirth" would seem to be an important factor to measure appropriately

** Yes. NQF staff note significant differences in mean performance between the chart measure and the eCQM measure.

2b2-3. Exclusions/Risk Adjustment

Comments:

** See 0469 comments

** no concerns

** The developer states: "There is no compelling evidence available supporting association between social risk factors and this measure." With our current knowledge of institutional racism and implicit bias, I would like to see a deeper analysis of the data around social risk factors.

2b4-7. Threats to Validity/Meaningful Differences/Comparability of Performance Scores/Missing Data

Comments:

** See 0469 comments

** Some concern about validity, especially given the data issues and the significant difference compared to the non-e measure

** The developer reports that missing rates for data elements is decreasing over time but has not been fully resolved.

Criterion 3. Feasibility

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

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

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

Question for the Committee:

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

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

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

Comments:

** See 0469 comments

** no concerns

** Although there are discrepancies in which data are available in paper and electronic formats, I think it is critical that we continue to improve electronic data collection systems. I support continuing to use this measure.

Criterion 4: Usability and Use

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

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

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

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

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

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

Accountability program details

- CMS's Hospital Inpatient Quality Reporting Program (payment)
- Joint Commission's Hospital Accreditation Program (regulatory/accreditation)
- Public reporting is planned

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

- Questions on the measures most likely come through the clinical and data receipt mailboxes provided on all communications. In addition, the Joint Commission has advisory committees for the Hospital Accreditation Program, which meet quarterly and have the opportunity to provide feedback.
- The Joint Commission utilizes an automated feedback system with access available to the measured entities and the vendors contracted by measured entities. A clinical lead is responsible for each individual measure set. The system is monitored daily and response is provided typically within eight business hours. If queries cannot be managed via written response, arrangements are made to address any issues or concerns via phone. All feedback is tracked and considered. If upon analysis there are trends noted giving cause for updates, this is reviewed by the measure work-group to confirm the need for revision.

- The Joint Commission engages a Technical Advisory Panel (TAP) for review and/or approval of updates which require additional subject matter expertise. All measure specifications are reviewed twice a year and updates are made as needed based on feedback from the measure users, input from the TAP, changes in the guidelines, or changes in clinical practice.
- Minor modifications have been made to this measure based upon feedback received.
- eCQM data are not being publicly reported by The Joint Commission.

Additional Feedback:

- Not reviewed by the Measure Applications Partnership

***Questions for the Committee**:*

- *Can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Does the Committee wish to discuss with the developer the timeline for public reporting of eCQM?*

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

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

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

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

Improvement results

- Performance data from the 2016 version were not available, so results cannot be compared to 2018 data to assess improvement
- Data for "paper" measure (NQF 0649) show improvement, however the developer did not summarize trends over time in the 2020 submission for the period 2015-2017. The 2018 mean rate was 1.7% (1,616 hospitals, 139,213 patients), which was an improvement from CY 2014 (3.3% national aggregate rate).
- The high performance on the chart measure raises the question as to whether it is topped out (and accordingly whether this measure is).

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

- No unexpected findings encountered during implementation

Potential harms

- None identified

Additional Feedback:

- None

***Questions for the Committee**:*

- *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**

RATIONALE:

- Performance data for 2015--> current were not provided, so usability cannot be assessed.

- Performance data for 2018 on the chart version may indicate the measure is "topped out" (mean 1.7%, SD 2.8%). Mean performance for 2018 on the eCQM measure was 17.6%, SD 21.6%, which the developer notes arises from data element validity/capture. Given the measures are to be publicly reported as part of Quality Check and are currently used in CMS accountability programs, does the Committee wish to discuss with the developer how results from the two measures are reconciled/interpreted?

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

4a1-2. Use - Accountability and Transparency/Feedback

Comments:

** See 0469 comments

** would like to know the timeline for public reporting and how this will be reconciled with the related measure

** This measure is not currently publicly reported.

4b1. Usability – Improvement/ Benefits vs. harms/ Transparency

Comments:

** See 0469 comments

** similarly, how will this be reconciled with the related measure

** It seems to me that accuracy of maternal health data has been a lower priority for EHR developers than that of other medical specialties. Until vendors are held accountable for accuracy of maternal health data, it will be a slow process to usability.

Criterion 5: Related and Competing Measures

Related or competing measures

- This measure is related to 0469, the paper medical record version.

Harmonization

- The developer states that the two measures are "completely harmonized to the extent possible," given the fact that the data source for #0469 is the paper medical record, and the data source for #2829 is the electronic health record.

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

Comments:

** See 0469 comments

** it seems like more work needs to be done to harmonize the related measures

** This measure is intended to be harmonized with #0469. However, additional work is required.

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of June 15, 2020

- The Federation of American Hospitals (FAH) appreciates the opportunity to comment on measure #469e, PC-01 Elective Delivery, prior to the Standing Committee's evaluation. Specifically, the FAH asks the committee to discuss potential concerns with the validity of this electronic clinical quality measure (eCQM) in light of the kappa scores for two of the data elements (medical induction of labor and active labor). Because these data elements are integral to calculating the performance of the eCQM, the FAH does not believe that this measure meets the validity subcriterion and thus may not be appropriate for accountability uses at this time.
- **Of the 1 NQF members who have submitted a support/non-support choice:**
    - 0 support the measure
    - 1 do not support the measure

Developer Submission

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

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

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

2020_nqf_evidence_attachment_ePC01_0469e_.docx

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

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

Yes

1a. Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0469e

Measure Title: ePC-01 Elective Delivery

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: April 8, 2020

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

☐ Outcome: Click here to name the health outcome

☐ Patient-reported outcome (PRO): Click here to name the PRO

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health- related behaviors. (*A PRO-based performance measure is not a survey instrument. Data may be collected using a survey instrument to construct a PRO measure.)*

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

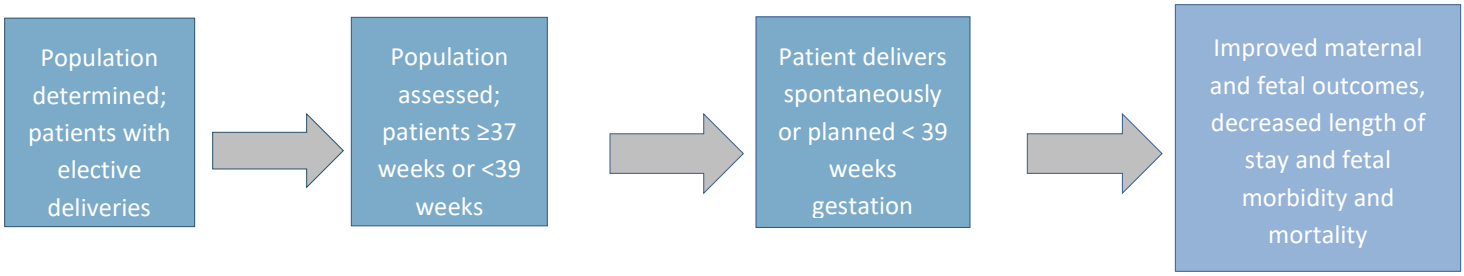
☐ Process: Patients with elective vaginal deliveries or elective cesarean births at >= 37 and < 39 weeks of gestation completed

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

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

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



The intent of the measure is to decrease the number of elective deliveries before 39 weeks>> population determined; patients with elective deliveries >> population assessed; patients ≥37 weeks and < 39 weeks>> patient delivers spontaneously or planned delivery <39 weeks gestation >> improved maternal and fetal outcomes; decreased length of stay and fetal morbidity and mortality.

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

Not applicable

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

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

Not applicable

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

1a.4) If you wish to include more than one systematic review, add additional tables.

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

☐ Clinical Practice Guideline recommendation (publication in the table)

☐ US Preventive Services Task Force Recommendation

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

☐ Other See 1a.4

Source: • Title • Author • Date • Citation, including page number • URL	**Title:** Avoidance of nonmedically indicated early-term deliveries and associated neonatal morbidities. **Author:** Borders, E.B., Birsner, M.L., Gyanmfi-Bannerbaum, C. **Date:** 2019 **Citation:** Borders, E.B., Birsner, M.L., Gyanmfi-Bannerbaum, C. (2019). Avoidance of nonmedically indicated early-term deliveries and associated neonatal morbidities. *American College of Obstetricians and Gynecologists Committee Opinion, 133:2,* e156-163. **URL:** https://www.acog.org/-/media/Committee-Opinions/Committee-on-Obstetric-Practice/co765.pdf?dmc=1&ts=20191119T1727071568
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	**From guideline abstract:** A recent systematic review found that late-preterm and early-term children have lower performance scores across a range of cognitive and educational measures compared with their full-term peers. Because non-respiratory morbidities also are increased in early-term deliveries, documentation of fetal pulmonary maturity does not justify an early nonmedically indicated delivery. Amniocentesis for the determination of fetal lung maturity should not be used to guide the timing of delivery, even in sub optimally dated pregnancies. The American College of Obstetricians and Gynecologists (ACOG) and the Society for Maternal–Fetal Medicine make the following recommendations: 1. Nonmedically indicated delivery, including cesarean delivery, inductions of labor, and cervical ripening should not occur before 39 0/7 weeks of gestation. 2. Implementation of a policy to decrease the rate of nonmedically indicated deliveries before 39 0/7 weeks of gestation has been found to decrease the number of these deliveries and, as a result, improve overall neonatal outcomes. 3. Avoidance of a nonmedically indicated delivery before 39 0/7 weeks of gestation is distinct from, and should not result in, an increase in expectant management of patients with medical indications for delivery before 39 0/7 weeks of gestation. 4. Indications for delivery before 39 0/7 weeks of

	gestation should be documented clearly and discussed with the patient. 5. Because nonrespiratory morbidities also are increased in early-term deliveries, documentation of fetal pulmonary maturity does not justify an early nonmedically indicated delivery. Amniocentesis for the determination of fetal lung maturity should not be used to guide the timing of delivery, even in sub optimally dated pregnancies.
Grade assigned to the evidence associated with the recommendation with the definition of the grade	No grades of evidence were assigned to the recommendations.
Provide all other grades and definitions from the evidence grading system	Not applicable
Grade assigned to the recommendation with definition of the grade	Not applicable
Provide all other grades and definitions from the recommendation grading system	Not applicable
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	**Quantity:** This committee opinion by the American College of Obstetricians and Gynecologists has identified 24 studies that helped to support the evidence in reducing early elective deliveries as presented by the committee. Of those studies, there were retrospective cohort studies (11), cohort studies (8), observational studies (2), sequential ecological studies (2), population-based prospective cohort analysis (1). **Quality:** Information on the overall quality of evidence across the studies is not provided; although, this committee opinion discusses the evidence supporting the reduction of elective deliveries >= 37 and < 39 weeks of gestation. As states have effectively reduced early elective delivery, multiple studies using national population level data have shown that even as the gestational age at term has increased in response to efforts to reduce early elective delivery, these efforts have not adversely affected stillbirth rates nationally or even in states with the greatest reductions in early elective delivery. There is no documented evidence regarding controversy related to the reduction of non-medically indicated early term elective deliveries. A review of studies also supports the use of quality improvement interventions to further reduce the number of such deliveries. **Quantity:** High **Quality:** Moderate **Consistency:** Moderate

Estimates of benefit and consistency across studies	Estimates of benefit and consistency across the studies are not provided; although, this committee opinion discusses the evidence supporting the reduction of elective deliveries >= 37 and < 39 weeks of gestation. **Benefits:** The central topic for the measure is to reduce elective deliveries >= 37 and < 39 weeks of gestation completed. The target population for the performance measure is consistent with the body of evidence supporting the reduction of elective deliveries. Implementation of a policy to decrease the rate of nonmedically indicated deliveries before 39 0/7 weeks of gestation has been found to decrease the number of these deliveries and, as a result, improve overall neonatal outcomes. A recent study examined the implementation of three approaches to this issue: 1) a hard-stop policy, which prohibited nonmedically indicated deliveries at the hospital level; 2) a soft-stop policy, in which health care providers agreed not to perform nonmedically indicated deliveries before 39 weeks of gestation; and 3) an education program that informed health care providers about the risks associated with delivery before 39 weeks of gestation. Overall, these approaches contributed to a greater than 50% reduction in the rate of nonmedically indicated early-term deliveries, regardless of the policy used (28). However, the reduction was the greatest in the hard-stop policy group, with a reduction from 8.2% to 1.7% (P=.007). The reduction was slightly less in the soft-stop policy group, with a reduction from 8.4% to 3.3% (P=.025), and the least in the educational approach group, with a reduction from 10.9% to 6.0% (P=.135), which was not statistically significant. **Consistency:** These studies demonstrate that a reduction in nonmedically indicated early-term and late-preterm deliveries can be achieved. Studies clearly have shown short-term and long-term outcomes are improved for infants born at full term (39 0/7–40 6/7 weeks of gestation) versus late preterm (34 0/7–36 6/7 weeks of gestation) or early term (37 0/7–38 6/7 weeks of gestation). The studies referenced in the committee opinion show a strong support of evidence that early elective deliveries pose a greater risk of unfavorable neonatal outcomes.
What harms were identified?	There have been no harms identified as a result of the implementation of the studies. However, the following

	harms listed below may occur as a result of early elective delivery. Neonatal Morbidities Associated with Early-Term Delivery: • Respiratory distress syndrome • Transient tachypnea of the newborn • Ventilator use • Pneumonia • Respiratory failure • Neonatal intensive care unit admission • Hypoglycemia • 5-minute Apgar score less than 7 • Neonatal mortality In a large cohort of planned term deliveries (defined as deliveries not initiated by labor or ruptured membranes) during a 3-month period in 27 hospitals across the United States, neonatal intensive care unit (NICU) admission rates were higher among neonates delivered in the early-term period. A comparison of NICU admission rates for neonates delivered at 37 weeks of gestation or 38 weeks of gestation with those for neonates delivered at 39 weeks of gestation revealed that 31% of 17,794 deliveries had no medical indication. Admission to the NICU, which can be dependent on a variety of factors, was required for 17.8% of infants delivered without medical indication at 37 weeks of gestation and for 8% delivered without medical indication at 38 weeks of gestation, compared with 4.6% of infants delivered at 39 weeks of gestation or beyond (P<.001 for deliveries at 38 weeks of gestation and 39 weeks of gestation). Harms identified with early term delivery are listed above.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	As a result of the literature search there have been no new studies conducted since this publication that would change the conclusions from the referenced recommendations.
2016 submission	
Source of Systematic Review: • Title • Author • Date • Citation, including page number • URL	American College of Obstetricians and Gynecologists Practice Bulletin No 107. (2009, August). Induction of labor. Obstetrics & Gynecology, 386-397. Retrieved from https://www.mnhospitals.org/Portals/0/Documents/patientsafety/Perinatal/acog--practice_bulletin_107_2009.pdf The American College of Obstetricians and Gynecologists the nation´s leading group of professionals providing health care for women. Practice Bulletins provide obstetricians and gynecologists with current information on established techniques and clinical management

	guidelines. The American College of Obstetricians and Gynecologists (the College) continuously surveys the field for advances to be incorporated in these series and monitors existing bulletins to ensure they are current. Individual bulletins are withdrawn from and added to the series on a continuing basis and reaffirmed periodically.
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	Clinical Considerations and Recommendations Page 389 What are the indications and contraindications to induction of labor? Indications for induction of labor are not absolute but should take into account maternal and fetal conditions, gestational age, cervical status, and other factors. Following are examples of maternal or fetal conditions that may be indications for induction of labor: • Abruptio placentae • Chorioamnionitis • Fetal demise • Gestational hypertension • Preeclampsia, eclampsia • Premature rupture of membranes • Post term pregnancy • Maternal medical conditions (e.g., diabetes mellitus, renal disease, chronic pulmonary disease, chronic hypertension, antiphospholipid syndrome) • Fetal compromise (e.g., severe fetal growth restriction, isoimmunization, oligohydramnios) Labor also may be induced for logistic reasons, for example, risk of rapid labor, distance from hospital, or psychosocial indications. In such circumstances, at least one of the gestational age criteria in the box should be met, or fetal lung maturity should be established. A mature fetal lung test result before 39 weeks of gestation, in the absence of appropriate clinical circumstances, is not an indication for delivery. The individual patient and clinical situation should be considered in determining when induction of labor is contraindicated. Generally, the contraindications to labor induction are the same as those for spontaneous labor and vaginal delivery. They include, but are not limited to, the following situations: • Vasa previa or complete placenta previa • Transverse fetal lie • Umbilical cord prolapse • Previous classical cesarean delivery

	• Active genital herpes infection • Previous myomectomy entering the endometrial cavity What criteria should be met before the cervix is ripened or labor is induced? Assessment of gestational age and consideration of any potential risks to the mother or fetus are of paramount importance for appropriate evaluation and counseling before initiating cervical ripening or labor induction. The patient should be counseled regarding the indications for induction, the agents and methods of labor stimulation, and the possible need for repeat induction or cesarean delivery. Although prospective studies are limited in evaluating the benefits of elective induction of labor, nulliparous women undergoing induction of labor with unfavorable cervices should be counseled about a twofold increased risk of cesarean delivery (Level II-2). In addition, labor progression differs significantly for women with an elective induction of labor compared with women who have spontaneous onset of labor (Level II-2). Allowing at least 12–18 hours of latent labor before diagnosing a failed induction may reduce the risk of cesarean delivery (Level II-2, 3). Additional requirements for cervical ripening and induction of labor include assessment of the cervix, pelvis, fetal size, and presentation. Monitoring FHR and uterine contractions is recommended as for any high-risk patient in active labor. Although trained nursing personnel can monitor labor induction, a physician capable of performing a cesarean delivery should be readily available. The American College of Obstetricians and Gynecologists the nation´s leading group of professionals providing health care for women. Practice Bulletins provide obstetricians and gynecologists with current information on established techniques and clinical management guidelines. The American College of Obstetricians and Gynecologists (the College) continuously surveys the field for advances to be incorporated in these series and monitors existing bulletins to ensure they are current. Individual bulletins are withdrawn from and added to the series on a continuing basis and reaffirmed periodically.
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	Although grading of the evidence was not determined during our systematic review, it was determined that the guideline developers accounted for a balanced representation of information, looked beyond one specialty group or discipline, and provided information that was accessible and met the requirements set out in this measure maintenance form.
Provide all other grades and definitions from the evidence grading system	Not applicable

Grade assigned to the **recommendation** with definition of the grade	Yes The American College of Obstetricians and Gynecologists Level II
Provide all other grades and definitions from the recommendation grading system	USPSTF
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	The central topic for the measure is the reduction of elective deliveries at >= 37 and < 39 weeks of gestation completed. The target population for the performance measure is consistent with the body of evidence supporting the reduction of elective deliveries. Quantity- No randomized-control trials (RCTs) were identified for early-term elective deliveries. RCTs were only identified for post-term elective deliveries versus expectant management. Given the current amount of population data available on the harms of early term and late pre-term delivery, it would be unethical to conduct such a study. Several studies were identified which were retrospective cohort or prospective observational in design examining thousands of births and the potential for adverse outcomes for both mother and newborn. In addition, several recent studies were identified addressing quality improvement interventions that were successful in reducing non-medically indicated early term elective deliveries. Quality- The quality of evidence supporting the reduction in the number of non-medically indicated elective deliveries is moderate. It is noteworthy to examine the fact that randomized control trials cannot be conducted, as one cannot randomly select women to agree to an elective delivery at < 39 weeks gestation. As previously noted, both ACOG and AAP have had guidelines in place for a number of years which do not support non-medically indicated elective deliveries at > 39 weeks gestation. Several studies consistently document increased morbidity associated with elective delivery before 39 weeks. The studies note that elective deliveries performed at < 39 weeks carry significant risk for the newborn (odds ratios 2.0-3-0 compared to newborns born between 39-41 weeks). In spite of the fact that all studies reviewed were either retrospective or prospective cohort studies, no study design flaws were noted. There is no documented evidence regarding controversy related to the reduction of non-medically indicated early term elective deliveries. A review of recent studies also supports the use of quality improvement interventions to further reduce the number of such deliveries. **Quantity: High**

	Quality: Moderate **Consistency: High**
Estimates of benefit and consistency across studies	The body of evidence consistently supports the benefit of reduction of non-medically indicated early term elective deliveries. All studies show an increase in the number of neonatal morbidities associated with early term deliveries, subsequent reduction of elective non-medically indicated deliveries reduces harm to the neonate. All studies demonstrated similar findings related to the direction of effect, though the magnitude varied from study to study, i.e., 8-17.8% increase in NICU admissions, rates of adverse respiratory outcomes, mechanical ventilation, newborn sepsis, hypoglycemia, admission to the NICU and hospitalization of 5 days or more increased by a factor of 1.8 to 4.2. and the incidence of transient tachypnea of the newborn, respiratory distress syndrome (RDS) and persistent pulmonary hypertension of the newborn were 3.1%, 0.25% and .17% respectively. As described before, elective deliveries performed at =>39 weeks gestation results in improved maternal and neonatal outcomes and will result in a substantial decrease in cesarean sections and neonatal morbidity, as well as substantial savings in health care costs. A recent study showed that by waiting until 39 weeks gestation, the NICU admissions fell from 12.8% to 5.9%, RDS fell from 3.7% to 0.9%, newborn sepsis fell from 7.0% to 2.5% and hospitalization > 5 days fell from 9.1% to 3.6%. This same study estimated that one-half million newborn intensive care unit days could be avoided in the U.S. population were a national rate of 1.7% to be achieved, with cost savings approaching $1 billion annually. There is no documented evidence regarding controversy related to the reduction of non-medically indicated early term elective deliveries. A review of recent studies also supports the use of quality improvement interventions to further reduce the number of such deliveries.
What harms were identified?	Not applicable
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	Not applicable

1a.4 OTHER SOURCE OF EVIDENCE

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

Not applicable

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

Not applicable

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

Not applicable

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

Not applicable

From previous submission: Citations from Evidence Other Than Guidelines

- American Academy of Family Physicians. (2000). Tips from Other Journals: Elective induction doubles cesarean delivery rate, 61, 4. Retrieved September 16, 2011 at: http://www.aafp.org/afp/20000215/tips/39.html.
- American College of Obstetricians and Gynecologists. (November 1996). ACOG Educational Bulletin.
- American College of Obstetricians and Gynecologists (ACOG) Practice Bulletin No 107. (2009). Induction of labor. Obstetrics & Gynecology. 114(2). 386-97.
- Clark, S., Miller, D., Belfort, M., Dildy, G., Frye, D., & Meyers, J. (2009). Neonatal and maternal outcomes associated with elective delivery. [Electronic Version]. Am J Obstet Gynecol. 200:156. e1-156.e4.
- Clark, S., Frye, D., Meyers, J., Belfort, M., Dildy, G., Kofford, S et al. (2010). Reduction in elective delivery at <39 weeks of gestation: comparative effectiveness of 3 approaches to change and the impact on neonatal intensive care admission and stillbirth. Am J Obstet Gynecol. 203:449. e1-6.
- Davidoff, M., Dias, T., Damus, K., Russell, R., Bettegowda, V.R., Dolan, S., et al. (2006). Changes in the gestational age distribution among U.S. singleton births; impacts on rates of late preterm birth, 1992-2002. Semin Perinatol. Feb;30(1):8-15.
- Engle, W.A. & Kominiarek, M.A. (2008). Late preterm infants, early term infants, and timing of elective deliveries. Clin Perinatol. 35:325-41.
- Glantz, J. (Apr.2005). Elective induction vs. spontaneous labor associations and outcomes. [Electronic Version]. J Reprod Med. 50(4):235-40.
- Tita, A., Landon, M., Spong, C., Lai, Y., Leveno, K., Varner, M, et al. (2009). Timing of elective repeat cesarean delivery at term and neonatal outcomes. [Electronic Version]. NEJM. 360:2, 111-120.

1b. Performance Gap

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

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

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

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

For almost 3 decades, the American College of Obstetricians and Gynecologists (ACOG) and the American Academy of Pediatrics (AAP established and followed the standard requiring 39 completed weeks gestation

prior to ELECTIVE delivery, either vaginal or operative (ACOG, 1996). A survey conducted in 2007 of almost 20,000 births, in HCA hospitals throughout the U.S. carried out in conjunction with the March of Dimes at the request of ACOG, revealed that almost one-third of all babies delivered in the United States are electively delivered with 5% of all deliveries in the U.S. delivered in a manner violating ACOG/AAP guidelines. Most of these are for convenience and result in significant short-term neonatal morbidity, such as neonatal intensive care unit admission rates of 13- 21% (Clark et al., 2009). A reduction in the number of non-medically indicated elective deliveries at >=37 to <39 weeks gestation will result in a substantial decrease in neonatal morbidity and mortality, as well as a significant savings in health care costs. In addition, the rate of cesarean deliveries should decrease with fewer elective inductions resulting in decreased length of stay and health care costs.

The measure will assist health care organizations (HCOs) to track non-medically indicated early term elective deliveries and reduce the occurrence.

Sources

- According to Glantz (2005), compared to spontaneous labor, elective inductions result in more cesarean births and longer maternal length of stay. The American Academy of Family Physicians (2000) also notes that elective induction doubles the cesarean delivery rate. Repeat elective cesarean births before 39 weeks gestation also result in higher rates of adverse respiratory outcomes, mechanical ventilation, sepsis and hypoglycemia for the newborns (Tita et al., 2009).
- American College of Obstetricians and Gynecologists. (November 1996). ACOG Educational Bulletin.
- American Academy of Family Physicians. (2000). Tips from Other Journals: Elective induction doubles cesarean delivery rate, 61, 4.Retrieved December 29, 2008 at: http://www.aafp.org/afp/20000215/tips/39.htm
- Clark, S., Miller, D., Belfort, M., Dildy, G., Frye, D., & Meyers, J. (2009). Neonatal and maternal outcomes associated with elective delivery. [Electronic Version]. Am J Obstet Gynecol. 200:156.e1-156.e4.
- Glantz, J. (Apr.2005). Elective induction vs. spontaneous labor associations and outcomes. [Electronic Version]. J Reprod Med. 50(4):235-40.
- Tita, A., Landon, M., Spong, C., Lai, Y., Leveno, K., Varner, M, et al. (2009). Timing of elective repeat cesarean delivery at term and neonatal outcomes. [Electronic Version]. NEJM. 360:2, 111-120.

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

Data are summarized at the hospital level.

Table 1. Median denominator size for Elective Delivery, 2018 (three-month reporting period, patients included in sample=14,181)

Number of Hospitals-131

Median number of deliveries-188

Median number of denominator cases-16

ePC-01 Distribution of Rates

2018 Data:

Rates on this measure: N=131, Mean 17.6%, SD 21.6%

10th Percentile= 0%

25th Percentile= 0%

50th Percentile= 13.3%

75th Percentile= 25.0%

90th Percentile= 41.4%

2016 Submission

This measure is a legacy eCQM that is currently included in the Hospital Inpatient Quality Reporting Program (HIQR) and the EHR Incentive Program. At present, no performance data are yet available.

In CY2016, CMS is requiring organizations participating in HIQR to electronically submit 1 quarter of data on 4 of 28 available eCQMs. This measure is one of the 28.

For more information, refer to CY2016 IPPS Final Rule, located here: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/FY2016-IPPS-Final-Rule-Home-Page.html

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

Not applicable

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

Based on 2018 discharges:

Measure rates by age category

Age	Rate (%)
<20	14.6
20-24	18.1
25-29	18.7
30-34	22.5
35-39	18.8
40+	32.7

Measure rates by Hispanic Ethnicity

Hispanic Ethnicity	Rate (%)
No	23.3
Yes	20.8

Measure rates by race

Race	Rate (%)
White	22.3
African American	18.3
American Indian	41.7
Asian	16.2
Pacific Islander	25.0
Other	15.0

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

2. Reliability and Validity—Scientific Acceptability of Measure Properties

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

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

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

Perinatal Health

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

Safety : Complications, Safety : Overuse

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

Women

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

https://www.jointcommission.org/specifications_for_joint_commission__electronic_clinical_quality_measures_ecqms.aspx

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

This is an eMeasure **Attachment:** ePC01_ElectiveDelivery_v8.zip

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

Attachment **Attachment:** PC01eCQM_Value_Sets_ReportingPeriod2020.xlsx

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

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

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

Not an instrument-based measure

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

Yes

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

On an annual basis, the eCQMs maintained by The Joint Commission undergo an annual update to revise specifications based on updated research and clinical information or standards changes. Changes have been made to the eCQM specifications in order to improve alignment with the chart abstracted measure from which this measure is derived. In 2017, the value set, Conditions Possibly Justifying Elective Delivery, was updated to include concepts representing Stillbirth and History of Stillbirth to better align with the chart abstracted measure. Due to the QDM 4.3 changes, the datatype was replaced for 'Physical Exam, Performed' to 'Assessment, Performed' for the data elements of Gestational Age, Time of Delivery, and Labor. Also, the timing logic for 'Time of Delivery' was updated to: < 1 day(s) starts before or concurrent with start of (´Occurrence A of Assessment, Performed: Time of Delivery´ starts during Occurrence A of EncounterInpatient) to better align with clinical intent. In 2018, two value sets for procedures performed were added to the history of prior uterine surgery procedure to better align with the clinical intent and the chart-abstracted version of the measure. Measure specifications were also converted from the Quality Data Model (QDM) measure logic to the Clinical Quality Language (CQL). In 2019, results from an empirical analysis comparing the disparities between the chart abstracted and eCQM data created the need to update the timing statement related to the documentation of the Estimated Gestational Age (EGA) to capture the last EGA assessment within one day or less prior to or at the same time of delivery.

Correction

The three data elements and value sets- 1.. "Diagnosis, Active: Complication of the Puerperium" using "Complication of the Puerperium ICD9CM Value Set (2.16.840.1.113883.3.117.1.7.1.375)" 2. "Diagnosis, Active: Complication Mainly Related to Pregnancy" using "Complication Mainly Related to Pregnancy ICD9CM Value Set (2.16.840.1.113883.3.117.1.7.1.372)" 3. Diagnosis, Active: Complication Mainly in the Course of Labor and Delivery" using "Complication Mainly in the Course of Labor and Delivery ICD9CM Value Set (2.16.840.1.113883.3.117.1.7.1.397)" were removed from the measure during the 2013 annual update and was based on coding guidance for the transition from ICD9 to ICD10. Feasibility was not reassessed since the measure was already in use with no change to the measure concept.

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

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

Inpatient hospitalizations for patients with elective deliveries by either:

- Medical induction of labor while not in labor prior to the procedure

- Cesarean birth while not in labor and with no history of a prior uterine surgery

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

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

The numerator includes the following two key items to calculate the cases from the target population.

-The 'Medical Induction' of labor should occur 24 hours or less before labor and is represented as a code from one of the following value sets and the associated QDM datatype:

o Procedure, Performed: Medical Induction of Labor (OID 2.16.840.1.113883.3.117.1.7.1.288)

o Procedure, Performed: Artificial Rupture of Membranes (OID 2.16.840.1.113762.1.4.1045.57)

o Medication, Administered: Oxytocin (OID 2.16.840.1.113762.1.4.1045.55)

o Medication, Administered: Dinoprostone (OID 2.16.840.1.113762.1.4.1045.56)

-The 'Labor' should occur during the delivery encounter and is represented with the QDM datatype and value set of Assessment, Performed: Labor (OID 2.16.840.1.113883.3.117.1.7.1.281)

- The 'Cesarean Birth' should start during the delivery encounter and is represented with the QDM data type and value set of "Procedure, Performed: Cesarean Birth (OID 2.16.840.1.113883.3.117.1.7.1.282)

- The history of 'Prior Uterine Surgery' should start before the start of the delivery encounter and is represented as a code from one of the following value sets and the associated QDM datatype:

o Diagnosis: Perforation of Uterus (OID 2.16.840.1.113762.1.4.1110.14)

o Diagnosis: Uterine Window (OID 2.16.840.1.113883.3.117.1.7.1.137)

o Diagnosis: Uterine Rupture (OID 2.16.840.1.113762.1.4.1110.16)

o Diagnosis: Cornual Ectopic Pregnancy (OID 2.16.840.1.113762.1.4.1110.12)

o Procedure, Performed: Classical Cesarean Birth (OID 2.16.840.1.113883.3.117.1.7.1.421)

o Procedure, Performed: Myomectomy (OID 2.16.840.1.113883.3.117.1.7.1.422)

o Procedure, Performed: Transabdominal Cerclage (OID 2.16.840.1.113762.1.4.1110.18)

o Procedure, Performed: Metroplasty (OID 2.16.840.1.113762.1.4.1110.25)

o Procedure, Performed: Uterine Horn (OID 2.16.840.1.113762.1.4.1110.24)

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

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

Inpatient hospitalizations for patients delivering newborns with >= 37 and < 39 weeks of gestation completed.

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

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

The denominator includes the following key elements:

- The delivery encounter must be less than or equal to 120 days during the measurement period and is represented with the QDM datatype and value set of Encounter, Performed: Encounter Inpatient (2.16.840.1.113883.3.666.5.307)

- The patient must be between the ages of 8 years and less than 65 years at the start of the delivery encounter and is represented with the QDM datatype and direct reference code of Patient Characteristic Birthdate: Birth date (LOINC Code 21112-8)

- The 'Delivery Procedure' should start during the delivery encounter and is represented with the QDM datatype and value set of Procedure, Performed: Delivery Procedures (OID:2.16.840.1.113762.1.4.1045.59)

- The 'Estimated Gestational Age' should be the last assessment within 1 day or less prior to or at the same time as the delivery and be greater than or equal to 37 weeks and less than 39 weeks and is represented with the QDM datatype and value set of Assessment, Performed: Estimated Gestational Age at Delivery (OID: 2.16.840.1.113762.1.4.1045.26)

- The 'Time of Delivery' should occur during the delivery encounter and is represented with the QDM datatype and value set of Assessment, Performed: Time of Delivery (OID: 2.16.840.1.113762.1.4.1045.28)

- The 'Time of Delivery' should occur during the delivery encounter and is represented with the QDM datatype and value set of Assessment, Performed: Time of Delivery (OID: 2.16.840.1.113762.1.4.1045.28)

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

Inpatient hospitalizations for patients with conditions possibly justifying elective delivery prior to 39 weeks gestation.

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

- The 'Conditions Possibly Justifying Elective Delivery' should be present during the delivery encounter and are represented with the QDM datatype, attribute and value set:

Diagnosis: Conditions Possibly Justifying Elective Delivery Prior to 39 Weeks Gestation using Conditions Possibly Justifying Elective Delivery Prior to 39 Weeks Gestation Grouping Value Set (2.16.840.1.113883.3.117.1.7.1.286)

Encounter diagnoses: Conditions Possibly Justifying Elective Delivery Prior to 39 Weeks Gestation using Conditions Possibly Justifying Elective Delivery Prior to 39 Weeks Gestation Grouping Value Set (2.16.840.1.113883.3.117.1.7.1.286)

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

Not Applicable, the measure is not stratified

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

No risk adjustment or risk stratification

If other:

S.12. Type of score:

Rate/proportion

If other:

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

Better quality = Lower score

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

See attached HQMF file.

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

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

Not Applicable.

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

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

Not Applicable. This measure is not based on a survey or a PRO-PM

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

If other, please describe in S.18.

Electronic Health Data, Electronic Health Records, Other

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

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

Hospitals report EHR data using Certified Electronic Health Record Technology (CEHRT), and by submitting Quality Reporting Document Architecture Category 1 (QRDA-1).

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

No data collection instrument provided

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

Facility, Other

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

Inpatient/Hospital

If other:

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

Not Applicable

2. Validity – See attached Measure Testing Submission Form

ePC01_Bonnie_ScreenShots.docx,2020_nqf_testing_attachment_ePC01_0469e_final-637227324523055268.docx

2.1 For maintenance of endorsement

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

Yes

2.2 For maintenance of endorsement

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

Yes

2.3 For maintenance of endorsement

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

No - This measure is not risk-adjusted

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

Measure Number (*if previously endorsed*)**:** 0469e

Measure Title: ePC-01 Elective Delivery

Date of Submission: January 3, 2020

Type of Measure:

□Outcome (*including PRO-PM*)	□Composite – ***STOP – use composite testing form***
☐Intermediate Clinical Outcome	□Cost/resource
□Process *(including Appropriate Use)*	□Efficiency
□Structure	

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

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

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

Measure Specified to Use Data From: (*must be consistent with data sources entered in S.17*)	**Measure Tested with Data From:**
□abstracted from paper record	□abstracted from paper record
□claims	□claims
□registry	□registry
□abstracted from electronic health record	□abstracted from electronic health record
□eMeasure (HQMF) implemented in EHRs	□eMeasure (HQMF) implemented in EHRs
□ other: Click here to describe	□ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*). Not applicable

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

2020 Submission

Testing of measure score reliability and validity was performed using data from hospital discharges occurring in 2018.

2016 Submission

The measure specifications were tested in the Bonnie testing environment which mimics the year 2012.

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

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.20*)	**Measure Tested at Level of:**
□individual clinician	□individual clinician

□group/practice	□group/practice
□hospital/facility/agency	□hospital/facility/agency
□health plan	□health plan
□ other: Click here to describe	□ other: Click here to describe

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

2020 Submission

This measure assesses the proportion of patients delivering newborns with >= 37 and < 39 weeks of gestation completed with an elective delivery. The intended use of the measure is to assess the quality of perinatal care in hospitals across the population.

Entities in reliability and validity testing: Results were calculated from Joint Commission data that included 131 hospitals submitting the measure using three months of 2018 discharges. These are records from hospitals that submitted both chart-abstracted and eCQM data for the same time period. The hospitals were geographically diverse and varied in size.

131 health care organizations representing various types, locations and sizes:
9 For Profit, 109 Not for Profit, 13 Government
40 >=300 beds; 54 100-300 beds; 37 <100 beds
37 Rural; 94 Urban
13 Major Teaching; 55 Minor Teaching; 63 Non-Teaching

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

2020 Submission

Patients in reliability and validity testing: Data are summarized at the hospital level. Below is a description of the sample. It includes number of hospitals included in Joint Commission data, the median initial population size, and the median denominator size for the measure across hospitals.

Table 1. Median denominator size for Elective Delivery, 2018 (three month reporting period, patients included in sample=14,181)

Number of Hospitals	Median number of deliveries	Median number of denominator cases
131	188	16

51 unique synthetic patient records were created in the BONNIE testing system for this measure. Cases were used to test the validity of each data element and timing relationship in the measure. Patient characteristics such as age, diagnosis, and length of stay were pre-determined to provide a variety of scenarios that adequately tested for patients passing each data element and failing each data element. Data included in cases and tested for this measure included diagnoses, differing types of delivery and uterine procedures, gestational ages, medications, patient orders, age, length of stay, and clinical observations, such as time of delivery.
All 51 cases passed or failed as expected based on the data included in the case, confirming the measure logic is accurate and valid.

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

2020 Submission

No differences in the data used for testing.

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

2020 Submission

No patient-level sociodemographic variables are used in the measure and none were available for analysis. There is no compelling evidence available supporting association between social risk factors and this measure.

2a2. RELIABILITY TESTING

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

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☐ **Performance measure score** (e.g., *signal-to-noise analysis*)

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

2020 Submission

Per NQF Measure Evaluation Criteria, reliability testing is not required if empiric validity of the data elements is assessed. See section 2b2 for validity testing of data elements.

The chart-abstracted version of this measure has been in national use since the 2nd quarter of 2010. At the time the chart-abstracted measure was originally tested, extensive tests of measure reliability were conducted. At present, no performance data for the electronic version of the measure are currently available. Below is the reliability testing summary for NQF #0469 PC-01 Elective Delivery Prior to 39 Completed Weeks Gestation, from which this measure is derived.

Demographics of organizations collecting and reporting data on these measures are as follows:
163 health care organizations representing various types, locations and sizes:
10 For Profit, 91 Not for Profit, 46 Military Facilities, 9 County, 2 State, 5 Other
15 >=500 beds; 29 250-499 beds; 50 100-249 beds; 69 <100 beds
Located in: AE, AK, AL, AP, AR, AZ, CA, DO, DC, FL, GA, HI, ID, IL, IN, KS, KY, LA, MA, MD, MI, MN, MO, MS, MT, NC, NE, NV, NY, OH,OK, PA, PR, RI, SC, TN, TX, VA, WA, WI, WV
26 performance measurement systems

Currently, hospitals are supported in their data collection and reporting efforts by 26 contracted performance measurement system (PMS) vendors. It is a contractual requirement of Joint Commission listed vendors that the quality and reliability of data submitted to them by contracted health care organizations must be monitored on a quarterly basis. In addition, The Joint Commission analyzes these data by running 17 quality

tests on the data submitted into ORYX. (ORYX is the term used by The Joint Commission to describe the component of the hospital accreditation program which requires data collection and reporting on standardized national performance measures). The following is a list of the major tests done on the submitted ORYX data, taken from the 2011 ORYX Performance Measurement System Requirements manual.

- Transmission of complete data
- Usage of individual core measure data received: To understand if the HCO provides the relevant service to treat the relevant population
- Investigation of aberrant data points
- Verification of patient population and sample size
- Identification of missing data elements
- Validation of the accuracy of target outliers
- Data integrity
- Data corrections

Data Element Agreement Rate:
Inter-rater reliability testing methodology utilized by contracted performance measure system vendors as outlined in the contract is as follows:

- All clinical data elements and all editable demographic elements are scored.
- All measure data are reabstracted with originally abstracted data having been blinded so that the reabstraction is not biased.
- Reabstracted data are compared with originally abstracted data on a data element by data element basis. A data element agreement rate is calculated. Clinical and demographic data are scored separately, and an overall agreement rate is computed.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)
Not applicable

Data element agreement rates for the chart-abstracted version of this measure were reported to The Joint Commission for 1Q11. This reflects the findings of 108 hospitals, comprising 13,279 records (100% sample). The following table delineates calculated agreement rates for individual data elements that are used to compute measure rates for PC-01.

Data Elements with a Mismatch	Total Numerator	Total Denominator	Rate
Active Labor	33	35	94.29%
Gestational Age	639	712	89.75%

These agreement rates are considered to be well within acceptable levels.

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)
Not applicable

Agreement rates for individual data elements tested for the chart-abstracted version of this measure were within acceptable levels. Once data are available for analysis, it is expected that reliability tests of the eCQM version of this measure will yield similar results.

2b1. VALIDITY TESTING
2b1.1. What level of validity testing was conducted? (*may be one or both levels*)
☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☐ **Performance measure score**

☐ **Empirical validity testing**

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

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

2020 Submission

Validity testing of critical data elements was performed by comparing eCQM data to the corresponding chart-based data that was submitted on the same patient. Patient-level data was matched to PC-01 chart-based data that was also transmitted to the Joint Commission (matched using hospital ID, admission date, discharge date and birth date) and each data element was compared between the ePC-01 data and the corresponding PC-01 data. Sensitivity, specificity and kappa statistics were used to measure the agreement between the two data sources, with the chart data considered to be the golden standard. The data elements birth date, admission date and discharge date used in the matching were assumed to be correct and could therefore not be used to measure agreement. The PC-01 chart-based data has been demonstrated to have high degree of data element reliability. The measure result was also compared between the two data sources and sensitivity, specificity and kappa statistics were used to measure agreement. The ePC-01 rates would be hypothesized to correlate positively with the PC-01 rates. In addition, the ePC-01 rate was correlated with other measures of perinatal care quality. Since a low measure rate for ePC-01 is desirable, this measure is hypothesized to correlate negatively to other perinatal care measures where a high rate is desirable (PC-05, ePC-05, PC-01) and positively to perinatal care measures where a low rate is desirable (PC-02).

A correlation of 0.1 - 0.3 was considered weak, 0.3 - 0.5 was considered moderate, and over 0.5 was considered strong. A kappa score above 0.6 was considered good and above 0.8 was considered excellent.

The Bonnie testing tool and environment were used to test the measure logic and value sets. Each data element and logic statement was tested to confirm actual results met expectations. Bonnie testing includes negative and positive testing of each data element in the measure. Positive testing ensures patients expected to be included in the measure are included. Negative testing ensures that patients who do not meet the data criteria are not included in the measure. An example of negative testing would be to include test cases with pediatric ages to ensure that pediatric patients are not included in the measure.

Denominator test cases positively test to ensure that only patients >=8 and <65 years of age who deliver an infant >=37 weeks and <39 weeks at time of delivery are included in the denominator. Negative test cases ensure that patients who do not meet these criteria to do not pass into the denominator. For example, cases test patient ages of 7, 8, 64, and 65 to ensure that patient age 7 and 65 do not pass into the denominator, but patients age 8 and 64 are included.

Numerator test cases positively test to ensure patients with elective deliveries by either medical induction of labor prior to the start of labor, or by cesarean birth while not in labor and with no history of prior uterine surgery are included in the numerator. Negative test cases ensure that a patient who did not meet these criteria are not included. For example, test cases in which labor began prior to cesarean birth ensure that patients who have a cesarean procedure as a result of complications of labor are not included in the numerator.

Denominator exclusion test cases for this measure ensure that patients are properly removed from the denominator if they have documented diagnoses that reflect conditions possibly justifying elective delivery.

Negative test cases for the denominator exclusion ensure that patients without these diagnoses fall in to the denominator population. Testing confirmed patients meeting the exclusion and exception criteria are removed from the measure appropriately, while those that do not meet the criteria are retained in the denominator population.

A review of the measure specifications was also conducted to confirm the logic was properly expressed within the current version of the QDM and confirmed the logic matches the clinical intent of the measure, as stated in the measure header. This is done through use of a logic checklist to facilitate review of the measure logic, according to several checks, a few examples of which are included below:

- Is the intent of the measure described in the measure description articulated/ captured in the measure logic?
- Do the logic elements map to definitions in the measure narrative, data dictionary or supporting reference documentation?
- Do the populations in the narrative align with the populations defined in the logic?
- Are the mathematic inequalities reflective of the measure intent and represent the intended populations (for example: when intended the inequality represents less than rather than less and or equal to)?

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

2020 Submission

Data Element Validity: Comparison of Electronic EHR extraction and manual chart abstraction

Measure Score Validity Statistics for Sample Between Electronic EHR Extraction and Manual Chart Abstraction (Sensitivity, Specificity, Kappa): Data Elements

Measure Component	N	Sensitivity	Specificity	Kappa (95% CI)
Gestational age in range	3738	97.8%	98.4%	.96 (.95, .97)
Conditions possibly justifying elective delivery	2859	96.1%	97.2%	.91 (.89, .93)
Cesarean birth	3753	94.3%	96.5%	.91 (.89, .92)
Medical induction of labor	3753	52.9%	76.4%	.30 (.27, .33)
Active labor	674	68.0%	32.4%	.00 (-.01, .08)
Prior Uterine Surgery	326	0%	100%	NA

Empirical Measure Score Validity:

Table 3. Measure Score Validity Statistics for Sample Between Electronic EHR Extraction and Manual Chart Abstraction (Sensitivity, Specificity, Kappa): Measure Score

Measure Component	Sensitivity	Specificity	Kappa (95% CI)
Initial patient population/denominator	88.0%	96.3%	.83 (.81, .85)
Numerator	44.4%	83.3%	.01 (-.01, .10)

Correlation with other measures of perinatal care quality

Measure	PC-01	PC-02	PC-05	ePC-01	ePC-05
PC-01-Elective Delivery	1				
PC-02-Cesarean Birth	0.133192	1			
PC-05-Exclusive Breast Milk Feeding	-0.02553	-0.28103	1		
ePC-01-Elective Delivery	0.008936	0.108322	0.022812		
ePC-05-Exclusive	0.040365	-0.17522	0.748033	-	1

Breast Milk Feeding				0.45737	

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

2020 Submission

Data Element Validity

Three of the most critical data elements show almost perfect agreement, with kappa scores above 0.90. The other data elements that are used later in the algorithm, and which impact relatively fewer cases, have fair to poor agreement due to the measure not being able to identify the presence of the data element, perhaps due to workflow and data capture issues in the eCQM. Overall, we believe the most critical data elements required for the measure show validity, however the fair to poor agreement on the other data elements has a large impact on the identification of numerator cases. In order to improve data capture, the Joint Commission provides an annual educational webinar series to discuss the logic interpretation, documentation workflow and data criteria for the electronic perinatal care measures.

Empirical Measure Score Validity

The determination of whether a case belongs in the measure population shows almost perfect agreement, with a kappa score above 0.80. However, the agreement of whether a case belongs in the numerator is poor, mainly due to the eCQM placing too many cases in the numerator. Since the total number of cases that the chart-based results places in the numerator was small, this results in the eCQM measure rates being higher than they should be. The inability of the eCQM to correctly identify active labor and prior uterine surgery has a major impact on the false identification of numerator cases. However, as experience is gained through the reporting of this measure, we expect this to improve over time. As this was based on 2018 data, changes have been made to the measure in 2019 to improve the data capture rate of gestational age; an empirical analysis comparing the disparities between the chart abstracted and eCQM data, the timing statement related to the documentation of the Estimated Gestational Age (EGA) was updated to capture the last EGA assessment within one day or less prior to or at the same time of delivery.

Correlation with other measures of perinatal care quality

Except for the correlation between ePC-01 and PC-05, the directions of the correlations were in the expected direction. The correlation of ePC-01 and PC-05, although expected to be in the negative direction, was not significantly greater than zero. The perinatal care measures used in this analysis are assessing different components of perinatal care and would not be expected to be more than weakly correlated since perinatal care quality is a multidimensional quantity.

_____

_____

2b2. EXCLUSIONS ANALYSIS

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

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

2020 Submission

Our testing addresses exclusions, as shown in the table below.

Measure Exclusions

Exclusion	Rationale
At least one condition possibly justifying elective delivery	This value set contains diagnosis codes and SNOMED codes for medical conditions that are reasons to perform an early term

	medical induction and/or cesarean delivery.
Gestational age < 37 or >=39	The denominator population is limited to patients > 37 to < 39 weeks of completed gestation. Patients with UTD for gestational age typically have had no prenatal care. Babies delivered at 37 to 38 6/7 weeks have a higher risk of complications than those delivered after 39 weeks. Therefore, these exclusions restrain the IPP (Initial Patient Population) to the population of interest.

We tested whether the exclusions affected overall performance score denominators.

Exclusions in the eCQM align with the chart-based version of the measure and are clinically necessary for the interpretation of the measure. As noted previously, the chart-abstracted version of this measure has been in national use since the 2nd quarter of 2010, and no data are available for the eCQM version of the measure. The below analysis addresses exclusions testing performed for the chart-abstracted version of the measure from which this measure is derived.

There were 1,134,640 admissions selected from the initial cohort. From among the 1,134,640 admissions in 1,237 hospitals, the descriptive statistics are given below.

The following exclusions were analyzed by subpopulation and measure for frequency and variability across providers:

- ICD-9-CM Principal Diagnosis Code or ICD-9-CM Other Diagnosis Codes for conditions possibly justifying elective delivery prior to 39 weeks gestation as defined in Appendix A, Table 11.07
- Less than 8 years of age
- Greater than or equal to 65 years of age
- Length of stay > 120 days
- Enrolled in clinical trials
- Gestational Age < 37 or >= 39 weeks or UTD

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

2020 Submission

Number and percent of denominator remaining after exclusions

ePC-01 denominator before exclusions	ePC-01 denominator after exclusions	Percent after exclusions
10,366	3,341	32.2%

The percentiles for the hospital percent after exclusions had the following values for the 10th, 25th, 50th, 75th and 90th percentiles respectively: 16.6%, 25.2%, 33.6%, 46.3%, and 66.7%.

Exclusion Subpopulation 1 - PC-01

ICD-9-CM Principal Diagnosis Code or ICD-9-CM Other Diagnosis Codes for conditions possibly justifying elective delivery prior to 39 weeks gestation as defined in Appendix A, Table 11.07

Exclusion: No observations noted

Less than 8 years of age

Exclusion: Included in the initial population exclusion

Greater than or equal to 65 years of age

Exclusion: Included in the initial population exclusion

Length of Stay >120 days
Exclusion: Included in the initial population exclusion

Exclusion: Enrolled in Clinical Trials
Overall Number of Occurrences n = 748
Overall Occurrence Percentage: 0.07%
Minimum: 0%
10th Percentile: 0%
Median: 0%
90th Percentile: 0.062%
Maximum: 28%

Exclusion: Gestational Age < 37 or gestational Age = >39 weeks or UTD
Overall Number of Occurrences n = 851,258
Overall Occurrence Percentage: 84.9%
Minimum 0.29%
10th Percentile: 69.17%
Median: 75.2%
90th Percentile: 79.2%
Maximum: 84.8%

2b2.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)
2020 Submission
We tested several exclusions in order to understand the impact on the denominator. Exclusions 1 and 2 are necessary to ensure the construct validity of the measure and all have a clinical rationale. The exclusions have an appreciable impact on those cases included in the denominator of the measure. In the specifications, these exclusions have been incorporated into the denominator definition.

Analysis of these data for the chart-abstracted progenitor of this measure indicated that all exclusions were appropriate. It is believed that results for exclusions in the eCQM will be similar when sufficient data have been received to perform such an analysis.

2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b4.

2b3.1. What method of controlling for differences in case mix is used?
☐ **No risk adjustment or stratification**
☐ **Statistical risk model with** Click here to enter number of factors **risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b3.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.
Not applicable

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

Not applicable

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

Not applicable

Also discuss any "ordering" of risk factor inclusion; for example, are social risk factors added after all clinical factors?

Not applicable

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

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

Not applicable

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

Not applicable

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

Not applicable

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

Not applicable

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.

If stratified, skip to 2b3.9

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

Not applicable

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

Not applicable

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

Not applicable

2b3.9. Results of Risk Stratification Analysis:

Not applicable

2b3.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)
Not applicable

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

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

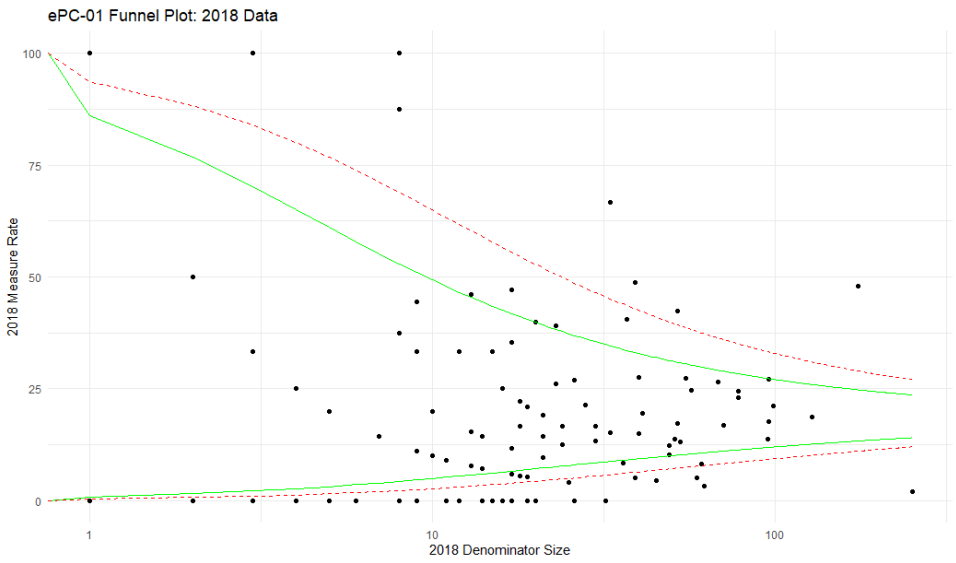
2020 Submission
To demonstrate meaningful differences in performance, The Joint Commission calculated a funnel plot (Spiegelhalter 2004) for the annual hospital rates of the measure. In a funnel plot, the observed measure is plotted against a measure of its precision, so that the control limits form a 'funnel' around the target outcome. The 95 per cent (≈2 standard deviation) and 99.8 per cent (≈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. Funnel plots for comparing institutional performance. Statistics in Medicine 2005; 24:1185–1202.

The method used to analyze meaningful differences in performance at The Joint Commission is Target Analysis. The object of target analysis is to compare a health care organization's (HCO) data against a comparative norm for the purpose of evaluating performance improvement opportunities. When an organization's performance level is statistically significantly different from a comparative norm, it is considered a statistical deviation. A statistical deviation may be desirable or undesirable depending on the "direction of improvement" of the measure.
There are two components to the target analysis methodology used at The Joint Commission. Given the national average for a performance measure, a target range is constructed. Using generalized linear mixed models' methodology (also known as hierarchical models), a predicted estimate of an HCO's performance, with a corresponding 95% confidence interval, is generated. This confidence interval is compared to the target range, to determine the HCOs' rating. The estimate of the organization's true performance is based on both the data from that organization and on data from the entire set of reporting organizations. A similar methodology will be used for the eCQMs.

2b4.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)
2020 Submission
Using the funnel plot, out of the 131 hospitals reporting, 14 hospitals were identified as high outliers with rates beyond the 2 standard deviation upper limit and 8 hospitals were identified as high outliers with rates beyond the 3 standard deviation limits.

Funnel Plot for ePC-01:



ePC-01 Distribution of Rates
2018 Data:
Scores on this measure: N=131, Mean 17.6%, SD 21.6%
10th Percentile= 0%
25th Percentile= 0%
50th Percentile= 13.3%
75th Percentile= 25.0%
90th Percentile= 41.4%

NQF: 0469: PC-01 Distribution of Outliers

2011 1st Quarter Data:
Scores on this measure: N=160, Mean 13.6%, SD 0.1594
10th Percentile= 0%
25th Percentile= 0%
50th Percentile= 9%
75th Percentile= 19%
90th Percentile= 34%

156 (97.5%) Neutral – results not significantly different from target range
4 (2.5%) Unfavorable - results statistically significantly lower than the national rate

2b4.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)
2020 Submission
The results indicate that there is significant hospital variability in rates, even after considering the small denominator sample sizes, and an appreciable number of hospitals that are not within the expected level of variability.

It should be noted that since data collection on this measure is completely voluntary for The Joint Commission, the hospitals reporting on this measure are self-selected, and therefore, presumably have a

particular interest and concern for improving perinatal care. For this reason, the measure results demonstrated by this group most likely significantly overstates the rate for the population of all health care organizations.

2b5. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
If only one set of specifications, this section can be skipped.

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

2020 Submission
This submission is for the eMeasure version of the measure 0469, which has been submitted as a separate measure.

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

2b5.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)
Not applicable

2b5.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)
Not applicable

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

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

2020 Submission
We used the data from the data element validity testing above to identify data elements that were missing in the eCQM but present in the chart. For each data element, a mixed model logistic regression was fit to the data, with the dependent variable being whether the data element was missing or not, and a chi-squared independence test was calculated to determine if there was significant between hospital variability in the missing data rates.

Data not present in the structured field from which the measure draws will not be included in the measure calculation. In the Bonnie testing environment, missing data are tested as an expected "Fail" for that data element, and actual performance (whether or not the case fails) is compared to expectations to ensure missing data impacts data element and overall measure calculation as expected. This is the extent of missing data analysis that can be performed in Bonnie.

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

2020 Submission

Missing data rates by data element (N=114 hospitals, Number of patients=14181)

Data element	min	25th %tile	mean	median	75th %tile	max	P-value independence test
Gestational age	0%	70.6%	77.9%	89.2%	100%	100%	< 0.0001
Conditions justifying elective delivery	0%	0%	28.1%	17.2%	59.9%	77.0 %	< 0.0001
Medical induction of labor	0%	2.4%	23.5%	32.2%	51.0%	72.2 %	< 0.0001
Cesarean birth	0%	0%	3.8%	0%	0.5%	100%	< 0.0001
Active labor	0%	1.1%	4.0%	2.3%	4.2%	70.0 %	< 0.0001
Prior uterine surgery	0%	0%	0.7%	0%	0.5%	20.0 %	< 0.0001

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

2020 Submission

There were significant differences in missing data rates across hospitals for all the data elements. Hospitals with 100% gestational age missing data rates would not have any cases falling in the measure and would therefore not have a measure rate to report. As this was based on 2018 data, changes have been made to the measure in 2019 to improve the data capture rate of gestational age; an empirical analysis comparing the disparities between the chart abstracted and eCQM data, the timing statement related to the documentation of the Estimated Gestational Age (EGA) was updated to capture the last EGA assessment within one day or less prior to or at the same time of delivery. For prior uterine surgery and active labor, missing data would result in cases being misidentified in the numerator with rates that are too high. We would expect the missing rate for these data elements to decrease with time as hospitals gain more experience with reporting this measure.

Hospital Data from 2017 were matched with the same hospitals for 2018. The distribution of rates is shown in the table below. On average, the rates are improving from 2017 to 2018.

	N	Mean	Std Dev.	Min	Q1	Median	Q3	Max
2017	53	0.2426	0.2785	0	0.0344	0.1538	0.2857	1
2018	53	0.2160	0.2196	0	0.0444	0.1731	0.3333	1

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

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

If other:

3b. Electronic Sources

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

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

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

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

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

Attachment: PC01_eCQM_NQF_Measure_Feasibility_Assessment_Report-635908844084639084.docx

3c. Data Collection Strategy

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

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

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

Upon analysis of the data received by the Joint Commission, there were differences in missing data rates across hospitals for all the data elements. Hospitals with 100% gestational age missing data rates would not have any cases falling in the measure. Therefore, these hospitals would not have a measure rate to report. As this was based on 2018 data, changes were made to the measure in 2019 to improve the data capture rate of gestational age. For the data element of prior uterine surgery and active labor, missing data would result in

cases being misidentified in the numerator with rates that are too high. The Joint Commission expects the missing rate for these data elements to decrease with time as hospitals gain more experience with reporting this measure, based upon continued analysis of data received and work with the accredited healthcare organizations on quality improvement related to the measure.

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

Value sets are housed in the Value Set Authority Center (VSAC), which is provided by the National Library of Medicine (NLM), in coordination with the Office of the National Coordinator for Health Information Technology and the Centers for Medicare & Medicaid Services.

Viewing or downloading value sets requires a free Unified Medical Language System® (UMLS) Metathesaurus License, due to usage restrictions on some of the codes included in the value sets. Individuals interested in accessing value set content can request a UMLS license at (https://uts.nlm.nih.gov/license.html)

There are no other fees or licensing requirements to use the Joint Commission performance measures, all of which are in the public domain.

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. Current and Planned Use

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

Specific Plan for Use	Current Use (for current use provide URL)
Public Reporting	Payment Program Hospital Inpatient Quality Reporting Program https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalRHQDAPU.html Regulatory and Accreditation Programs Hospital Accreditation Program http://jointcommission.org Hospital Accreditation Program http://jointcommission.org

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

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

Name of program and sponsor: Hospital Accreditation Program-The Joint Commission Hospital Accreditation Program

- 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 Nationwide; 3895 Joint Commission-accredited hospitals (2019)
- Geographic area and number and percentage of accountable entities and patients included Nationwide; 131 hospitals reported measure (4% of accredited hospitals) representing 14,181 patients (2018) Name of program and sponsor: The Joint Commission Perspective's: The Official Newsletter of the Joint Commission. (2019). The joint commission recognizes 20 years of ORYX performance measure reporting; look back at the 20-year evolution of performance measure reporting and review the ORYX chart-abstracted measure results for 2017 and 2018, 39, 10.
- Purpose: The Perspective's article provides authoritative, accurate, and timely information about revisions and updates to Joint Commission standards, policies, and other requirements for all Joint Commission-accredited and -certified organizations and healthcare settings.

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

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

Not Applicable.

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

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

The Joint Commission provides accredited healthcare organizations feedback reports for the measures submitted. The results are shared with organizations on a quarterly and/or annual basis depending on the reporting cycle of the measure. In addition, the Joint Commission has launched a new program called Continuous Customer Engagement (CCE) to assist organization in improving the quality of the performance measures. CCE includes enhanced dashboards with QI tools embedded into the dashboard, as well as focused and targeted solutions to assist organizations with gaps in the performance of their measures. The initial outreach to organizations utilizes an email process for hospital contact related to their measure rates and analysis. Response is provided in a timely manner either by email or directly by phone. Additionally, the data is available publicly through The Joint Commission Quality Check website. Individual hospital data for each rolling yearly time period is viewable and can be downloaded from this website. eCQM data is not being publicly reported by The Joint Commission.

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

The Joint Commission is committed to provided valuable and actionable feedback to accredited organizations submitted the performance measurement data. The Joint Commission aggregates the patient level data at the hospital level quarterly. The hospital Performance Measure Report and Quality Check website are updated either quarterly or annually to reflect organization results, as well as National Benchmarks. A user guide to the Performance Measure Report is posted on the Joint Commission website. Quality Check includes yearly and quarterly hospital rates, state and national averages, and the top 10 percentile at the national and state level.

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

Describe how feedback was obtained.

The Joint Commission utilizes an automated feedback system with access available to the measured entities and the vendors contracted by measured entities. A clinical lead is responsible for each individual measure set. The system is monitored daily and response is provided typically within 8 business hours. If queries cannot be managed via written response, arrangements are made to address any issues or concerns via phone. In addition, the Joint Commission developed dashboards as part of an ongoing project to provide continuous customer engagement. The Joint Commission analyzes aggregate performance in each of 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. 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 score-able element on 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.

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

The Joint Commission provides several venues for the organizations being measured to provide feedback. Questions on the measures are most likely to come through the clinical and data receipt mailboxes provided on all communications. In addition, the Joint Commission has advisory committees for the Hospital Accreditation Program, which meet on a quarterly basis, and have the opportunity to provide feedback on the measures being collected.

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

Same as above in 4a2.2.2

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

Note: All feedback is tracked and considered. If upon analysis there are trends noted giving cause for updates, this is reviewed by the measure work-group to confirm the need for revision. Additionally, The Joint Commission engages a Technical Advisory Panel (TAP) for review and/or approval of updates which require additional subject matter expertise. All measure specifications are reviewed twice a year and updates are made as needed based on feedback from the measure users, input from the TAP, changes in the guidelines, or changes in clinical practice.

Minor modifications have been made to this measure based upon feedback received.

Improvement

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

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

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

Not applicable

4b2. Unintended Consequences

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

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

We have not encountered any unexpected findings but continue to monitor feedback.

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

We have not encountered any unexpected findings but continue to monitor feedback.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

Yes

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

0469 : PC-01 Elective Delivery

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

Not applicable

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

5a.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):
Are the measure specifications harmonized to the extent possible?

Yes

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

The measures are completely harmonized to the extent possible, given the fact that the data source for #0469 is the paper medical record, and the data source for #2829 is the electronic health record.

5b. Competing Measures

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

OR

Multiple measures are justified.

5b.1. If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):

Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

Not Applicable.

Appendix

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

Available at measure-specific web page URL identified in S.1 **Attachment:**

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): The Joint Commission

Co.2 Point of Contact: JohnMarc, Alban, jalban@jointcommission.org, 630-792-5304-

Co.3 Measure Developer if different from Measure Steward: The Joint Commission

Co.4 Point of Contact: Tricia, Elliott, TElliott2@jointcommission.org, 630-792-5643-

Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

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

Michael Ross, MD, MPH (Chair)

Harbor-UCLA Medical Center

Torrance, CA

Martin McCaffrey, MD

UNC North Carolina Children's Hospital

Chapel Hill, NC

Debra Bingham PhD, RN, FAAN

Institute of Perinatal Quality Improvement

Washington, DC

James Christmas, MD

HCA Clinical Services Group

Elizabeth Rochin, PhD, RN, NE-BC

National Perinatal Information Center

Providence, RI

Cathy Ivory, PhD, RNC-OB, RN-BC, FAAN

Indiana University Health

Indianapolis, IN

Joseph Kunisch, PhD, RN-BC, CPHQ

Memorial Hermann Healthcare System

Houston, TX

B. Dale Magee, MD, MS

Shrewbury, MA

Elliott Main, MD

Stanford University

Mill Valley, CA

Susan Matney, PhD, RNC-OB

Intermountain Healthcare

Salt Lake City, UT

Elizabeth O'Neil-Greiner, RN, MHA

BJC Healthcare

St. Louis, MI

Patrick Romano, MD, MPH

University of California Davis Health

Sacramento, CA

Mark Tomlinson, MD

Providence Health System

Portland, OR

Brooke Villarreal, DNP, MSN, RN-BC

HCA Healthcare

Nashville, TN

The technical advisory panel (TAP) members determined priority areas that could be evaluated to improve care related to perinatal care during the development time frame. After implementation, minor revisions, acknowledged by TAP representatives, were made to improve clarity.

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2012

Ad.3 Month and Year of most recent revision: 07, 2019

Ad.4 What is your frequency for review/update of this measure? Annual

Ad.5 When is the next scheduled review/update for this measure? 06, 2020

Ad.6 Copyright statement: Measure specifications are in the Public Domain

Ad.7 Disclaimers: These performance measures are not clinical guidelines and do not establish a standard of medical care and have not been tested for all potential applications. The measures and specifications are provided without warranty.

Ad.8 Additional Information/Comments: