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 #: 0471

Corresponding Measures:

De.2. Measure Title: PC-02 Cesarean Birth

Co.1.1. Measure Steward: The Joint Commission

De.3. Brief Description of Measure: This measure assesses the rate of nulliparous women with a term, singleton baby in a vertex position delivered by cesarean birth. 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-02: Cesarean Birth is one of three measures in this set that have been re-engineered as eCQMs (ePC-01 Elective Delivery, ePC-02 Cesarean Birth and ePC-05 Exclusive Breast Milk Feeding).

A reduction in the number of nulliparous patients with live term singleton newborns in vertex position (NTSV) delivering by cesarean birth will result in increased patient safety, a substantial decrease in maternal and neonatal morbidity and substantial savings in health care costs, Main et al. (2011). Successful quality improvement efforts incorporate audit and feedback strategies combined with provider and nurse education, guidelines and peer review.

The measure will assist health care organizations (HCOs) to track nulliparous patients with live term singleton newborns in vertex position delivering by cesarean birth to reduce the occurrence. Nulliparous women have 4-6 times the cesarean birth rate than multiparous women; thus, the NTSV population is the largest driver of primary cesarean birth rate. Furthermore, nulliparity varies greatly among hospitals (20% to 60%) making it the most important risk factor for stratification or adjustment, Main et al. (2006). NTSV has the large variation among facilities, thus identifying an important population on which to focus quality improvement efforts.

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

Main, E.K., Moore, D., Farrell, B., Schimmel, L.D., Altman, R.J., Abrahams, C., et al., (2006). Is there a useful cesarean birth measure? Assessment of the nulliparous term singleton vertex cesarean birth rate as a tool for obstetric quality improvement. Am J Obstet Gynecol. 194:1644-51.

Main, E.K., Morton, C.H., Hopkins, D., Giuliani, G., Melsop, K. and Gould, J.B. (2011). Cesarean Deliveries, Outcomes, and Opportunities for Change in California: Toward a Public Agenda for Maternity Care Safety and Quality. Palo Alto, CA: CMQCC.

1b.1. Developer Rationale: The removal of any pressure to not perform a cesarean birth has led to a skyrocketing of hospital, state and national cesarean birth (CB) rates. Some hospitals now have CB rates over 50%. Hospitals with CB rates at 15-20% have infant outcomes that are just as good and better maternal outcomes (Gould et al., 2004). There are no data that higher rates improve any outcomes, yet the CB rates continue to rise. This measure seeks to focus attention on the most variable portion of the CB epidemic, the term labor CB in nulliparous women. This population segment accounts for the large majority of the variable portion of the CB rate, and is the area most affected by subjectivity.

As compared to other CB measures, what is different about NTSV CB rate (Low-risk Primary CB in first births) is that there are clear cut quality improvement activities that can be done to address the differences. Main et al. (2006) found that over 60% of the variation among hospitals can be attributed to first birth labor induction rates and first birth early labor admission rates. The results showed if labor was forced when the cervix was not ready the outcomes were poorer. Alfirevic et al. (2004) also showed that labor and delivery guidelines can make a difference in labor outcomes. Many authors have shown that physician factors, rather than patient characteristics or obstetric diagnoses are the major driver for the difference in rates within a hospital (Berkowitz, et al., 1989; Goyert et al., 1989; Luthy et al., 2003). The dramatic variation in NTSV rates seen in all populations studied is striking according to Menacker (2006). Hospitals within a state (Coonrod et al., 2008; California Office of Statewide Hospital Planning and Development [OSHPD], 2007) and physicians within a hospital (Main, 1999) have rates with a 3-5 fold variation.

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

The measure will assist health care organizations (HCOs) to track nulliparous patients with live term singleton newborns in vertex position delivering by cesarean birth to reduce the occurrence. Nulliparous women have 4-6 times the cesarean birth rate than multiparous women thus the NTSV population is the largest driver of primary cesarean birth rate. Furthermore, nulliparity varies greatly among hospitals (20% to 60%) making it the most important risk factor for stratification or adjustment. NTSV has the large variation among facilities thus identifying an important population on which to focus quality improvement efforts.

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

Sources

- Agency for Healthcare Research and Quality. (2002). AHRQ Quality Indicators Guide to Inpatient Quality Indicators: Quality of Care in Hospitals Volume, Mortality, and Utilization. Revision 4 (December 22, 2004). AHRQ Pub. No. 02-RO204.
- Alfirevic, Z., Edwards, G., & Platt, M.J. (2004). The impact of delivery suite guidelines on intrapartum care in “standard primigravida.” Eur J Obstet Gynecol Reprod Biol.115:28-31.
- American College of Obstetricians and Gynecologists. (2000). Task Force on Cesarean Delivery Rates. Evaluation of Cesarean Delivery. (Developed under the direction of the Task Force on Cesarean Delivery Rates, Roger K. Freeman, MD, Chair, Arnold W. Cohen, MD, Richard Depp III, MD, Fredric D. Frigoletto Jr, MD, Gary D.V.

Hankins, MD, Ellice Lieberman, MD, DrPH, M. Kathryn Menard, MD, David A. Nagey, MD, Carol W. Saffold, MD, Lisa Sams, RNC, MSN and ACOG Staff: Stanley Zinberg, MD, MS, Debra A. Hawks, MPH, and Elizabeth Steele).

- Bailit, J.L., Garrett, J.M., Miller, W.C., McMahon, M.J., & Cefalo, R.C. (2002). Hospital primary cesarean delivery rates and the risk of poor neonatal outcomes. Am J Obstet Gynecol. 187(3):721-7.
- Bailit, J. & Garrett, J. (2003). Comparison of risk-adjustment methodologies. Am J Obstet Gynecol.102:45-51.
- Bailit, J.L., Love, T.E., & Dawson, N.V. (2006). Quality of obstetric care and risk-adjusted primary cesarean delivery rates. Am J Obstet Gynecol.194:402.
- Bailit, J.L. (2007). Measuring the quality of inpatient obstetrical care. Ob Gyn Sur. 62:207-213.
- Berkowitz, G.S., Fiarman, G.S., Mojica, M.A., et al. (1989). Effect of physician characteristics on the cesarean birth rate. Am J Obstet Gynecol. 161:146-9.
- California Office of Statewide Hospital Planning and Development. (2017). Hospital Volume and Utilization Indicators for California, Retrieved from the Internet on February 22, 2018 at: https://www.oshpd.ca.gov/HID/AHRQ-Volume-Utilization.html
- Cleary, R., Beard, R.W., Chapple, J., Coles, J., Griffin, M., & Joffe, M. (1996). The standard primipara as a basis for inter-unit comparisons of maternity care. Br J Obstet Gynecol. 103:223-9.
- Coonrod, D.V., Drachman, D., Hobson, P., & Manriquez, M. (2008). Nulliparous term singleton vertex cesarean delivery rates: institutional and individual level predictors. Am J Obstet Gynecol. 694-696.
- DiGiuseppe, D.L., Aron, D.C., Payne, S.M., Snow, R.J., Dieker, L., & Rosenthal, G.E. (2001). Risk adjusting cesarean delivery rates: a comparison of hospital profiles based on medical record and birth certificate data. Health Serv Res.36:959-77.
- Gould, J., Danielson, B., Korst, L., Phibbs, R., Chance, K.,& Main, E.K., et al. (2004). Cesarean delivery rate and neonatal morbidity in a low-risk population. Am J Obstet Gynecol, 104:11-19.
- Goyert, G.L., Bottoms, F.S., Treadwell, M.C., et al. (1989). The physician factor in cesarean birth rates. N Engl J Med.320:706-9.
- Le Ray, C., Carayol, M., Zeitlin, J., Berat, G., & Goffinet, F. (2006). Level of perinatal care of the maternity unit and rate of cesarean in low-risk nulliparas. Am J Obstet Gynecol. 107:1269-77.
- Luthy, D.A., Malmgren, J.A., Zingheim, R.W., & Leininger, C.J. (2003). Physician contribution to a cesarean delivery risk model. Am J Obstet Gynecol.188:1579-85.
- Main, E.K. (1999). Reducing cesarean birth rates with data-driven quality improvement activities. Peds. 103: 374-383.
- Main E.K., Bloomfield, L., & Hunt, G. (2004). Development of a large-scale obstetric quality-improvement program that focused on the nulliparous patient at term. Am J Obstet Gynecol.190:1747-58.
- Main, E.K., Moore, D., Farrell, B., Schimmel, L.D., Altman, R.J., Abrahams, C., et al., (2006). Is there a useful cesarean birth measure? Assessment of the nulliparous term singleton vertex cesarean birth rate as a tool for obstetric quality improvement. Am J Obstet Gynecol. 194:1644-51.
- Menacker, F. (2005).Trends in cesarean rates for first births and repeat cesarean rates for low-risk women: United States, 1990-2003. Nat Vital Stat Rep. 54(4): 1-5.
- Romano, P.S., Yasmeen, S., Schembri, M.E., Keyzer, J.M., & Gilbert, W.M. (2005). Coding of perineal lacerations and other complications of obstetric care in hospital discharge data. Am J Obstet Gynecol.106:717-25.
- U.S. Department of Health and Human Services. (2000). Healthy People 2010: Understanding and Improving Health. 2nd ed. Washington, DC: U.S. Government Printing Office. Measure 16-9.

• Yasmeen, S., Romano, P.S., Schembri, M.E., Keyzer, J.M., & Gilbert, W.M. (2006). Accuracy of obstetric diagnoses and procedures in hospital discharge data. Am J Obstet Gynecol. 194:992-1001.

S.4. Numerator Statement: Patients with cesarean births with ICD-10-PCS Principal Procedure Code or ICD-10-PCS Other Procedure Codes for cesarean birth as defined in Appendix A, Table 11.06.

S.6. Denominator Statement: The outcome target population being measured is: Nulliparous patients with an ICD-10-CM Principal or Other Diagnosis Code for outcome of delivery as defined in Appendix A, Table 11.08 and with a delivery of a newborn with 37 weeks or more gestation completed or with an ICD-10-PCS Principal or Other Procedure Codes for delivery as defined in Appendix A, Tables 11.01.1.

S.8. Denominator Exclusions: • ICD-10-CM Principal Diagnosis Code or ICD-10-CM Other Diagnosis Codes for multiple gestations and other presentations as defined in Appendix A, Table 11.09

• Less than 8 years of age

• Greater than or equal to 65 years of age

• Length of Stay >120 days

• Gestational Age < 37 weeks or UTD

De.1. Measure Type: Outcome

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

S.20. Level of Analysis: Facility, Other

IF Endorsement Maintenance – Original Endorsement Date: Oct 24, 2008 **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 health outcome measure include providing empirical data that demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service; if these data not available, data demonstrating wide variation in performance, assuming the data are from a robust number of providers and results are not subject to systematic bias. For measures derived from patient report, evidence also should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.

Evidence Summary or **Summary of prior review in 2016**

- This outcome measure was last reviewed for maintenance in 2016, and the developer provides a logic model. Specifically, the measure's intent is to decrease the number of cesarean deliveries >> population determined in nulliparous, term singleton newborns in vertex position>> population assessed; nulliparous term patients>> patient delivers vaginally >> improved maternal and fetal outcomes; decreased length of stay and fetal morbidity and mortality.
- The 2016 evidence focused on a Level II recommendation from the 2000 monograph of the American College of Obstetricians and Gynecologists (ACOG) Task Force on Cesarean Delivery Rates. Evaluation of Cesarean Delivery.

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:

- The developer summarized updated information from two 2019 studies that support the ACOG recommendations.
  - The developer stated that the findings of Caughey, et al. support that reductions in cesarean delivery rates need not lead to worse neonatal or maternal outcomes.
  - The developer stated that the findings of Main et al., from a large-scale collaborative, provide evidence that a reduction in first birth cesarean delivery rates need not be associated with more difficult vaginal births or higher rates of major perineal lacerations. Further, the developer noted that the study found that the rate of severe unexpected newborn complications improved in hospitals with the greatest reduction in nulliparous, term, singleton, vertex cesarean delivery rates.

Question for the Committee:

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

Guidance from the Evidence Algorithm

Outcome measure (Box 1) → Relationship between outcome and at least one healthcare action is demonstrated by empirical data (Box 2) → Yes/Pass

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

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

Maintenance measures – increased emphasis on gap and variation

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

From the 2016 review:

- In January 2014, The Joint Commission required mandatory reporting of the perinatal care measure set for all accredited hospitals with 1,100 births or more annually; 1,388 hospitals reported the data with an average rate of 26.8% (n=363,400 patients). The 2014 performance gap persists with improvement noted primarily in the lower quartile (21.1%) and 10th percentile (17.6%) hospitals. It also noted that a performance gap of 12.4% exists for the 90th percentile of hospitals performing at 36.3% (if 23.9% is considered goal performance). The 2014 mean rate of 26.7% also remains above the HP 2020 goal.
- The threshold for mandatory reporting was lowered to 300 births annually effective January 2016. The new reporting requirement now captures approximately 80% of all accredited birthing hospitals.

For the 2020 submission:

The developer provided CY 2018 data, which it stated demonstrated considerable variability among hospitals, with over half of hospitals reporting rates above *The Healthy People 2020* goal of 23.9%.

- 1,936 hospitals, 490,481 patients:
  - Mean (SD): 25.7% (7.6%)
  - IQR: 8.9%
  - Deciles (0,10,20,30,40,50,60,70,80,90,100): 2.0%, 16.9%, 19.7%, 21.8%, 23.5%, 25.0%, 26.8%, 28.6%, 30.9%, 34.8%, 100%

The performance trend for this measure is as follows:

- 2015-26.2%
- 2016-26.1%
- 2017-26.0%
- 2018-25.5%

Disparities
Based on 2018 discharges:

- Performance by age category

Age	Rate (%)
<20	16.1
20-24	21.6
25-29	25.2
30-34	28.9
35-39	38.1
40+	53.0

- Performance by Hispanic Ethnicity

Hispanic Ethnicity	Rate (%)
No	24.9
Yes	24.7

- Performance by Race

Race	Rate (%)
White	24.1
African American	29.0
American Indian	24.3
Asian	25.0
Pacific Islander	26.7
Unable to Determine	24.2

Question for the Committee:

- Although overall, improvement/gap seems to be small on an aggregate basis, the developer provided data demonstrating more clearly disparities by age and race. Is there a gap and/or disparities in care that warrant a national performance measure?

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:

**Excellent evidence is reviewed in NQF measure worksheet -- new evidence further validates this measure.

**There is good evidence to support this measure and new evidence confirms previous findings.

**High evidence to support this measure. There continues to be a large gap in performance. Disparities are seen by age and race. This measure is a driver in the overall cesarean birth rate and is highly important to the target population.

**Evidence appears updated

**Arrive trial changes the previously held belief that inductions increase cesarean rates. However, while this study affects the logic behind the measure, there is still empiric evidence that using this measure to lower NTSV rates does not hurt babies or mother. I think what is changed is the presumption of how it can be lowered.

** strong evidence in support of this measure-

1b. Performance Gap

Comments:

**The data are very clear - there is a gap in care in terms of variability between birthing hospitals and overall less than optimal performance across the country.

**There are still significant gaps in care and this measure is still warranted.

**Large performance gap among delivering facilities and warrants a continued national focus.

**Yes, still significant variability and above stated target

**yes. wide range of NTSV rates over the percentiles. Unclear if the unadjusted gaps shown in age and race are reflective of a more complex patient population or a reflection on care. NTSV rates by race in particular, warrants further study.

** high

1b. Disparities

Comments:

**The most profound to me is a c-section rate of 29% in African American patients compared to 24.1% in white patients!

**Date show disparities in race and age.

**Disparities are seen between facilities and with age and race.

**Yes, race and age disparities. Would be interested in regional or insurance-status disparities

**see above

** significant disparities

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

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

Reliability

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

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

Validity

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

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

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

- This outcome measure was determined to not be complex and was not evaluated by the SMP.

Question for the Committee regarding reliability:

- *Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?*

Questions for the Committee regarding validity:

- *Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?*
- *The measure is not risk adjusted, but the developer provides a conceptual rationale and empirical analyses to justify this approach, specifically examining maternal age and BMI. Does the Committee wish to discuss the lack of risk adjustment with the developer?*

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

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

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 0471

Measure Title: PC-02 Cesarean Birth

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_0471" document, items S.1-S.22

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

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

   No concerns

RELIABILITY: TESTING

Submission document: "MIF_0471" 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

   - The developer provided score-level reliability for the measure using the beta-binomial model (signal to noise) on a data set of 1,936 hospitals with a median number of deliveries of 1,091; the median number of denominator cases was 142. Per Adams (2009), reliability scores range from 0.0 to 1.0. A score of zero implies that all variation is attributed to measurement error (i.e., noise), whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance (across hospitals).
   - Because of the switch from ICD-9 to ICD-10 codes, the developer also examined reliability by comparing results from the 2017 data submission that utilized ICD-9 codes to the results from the 2018 data submission that utilized ICD-10 codes. Summary statistics for the number of numerator cases, number of denominator cases, and observed rates are presented. Hospital data were also matched by each year in each of the three attributes and a paired t-test was used to determine statistical significance for each attribute.

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

   Submission document: Testing attachment, section 2a2.3

   - For the 2020 submission, the developer reports the following signal to noise reliability statistics:
     - Average: 0.76

- Median: 0.76
- 10th-90th percentile across hospitals: 0.63-0.92

- The developer noted that average score 0.76 is acceptable reliability for most of the hospitals. (As noted by the developer, in general, a score of 0.7 or higher suggests the measure has adequate reliability.)
- For the 2020 submission, the developer reports the following with respect to ICD-9 vs. ICD-10 codes:

Comparison between the ICD-9 and ICD-10 number of numerator cases, denominator cases, and observed rates

		N	Mean	Std. Dev.	Min	Q1	Median	Q3	Max	Pairwise Difference	P-Value
Numerator	ICD-9	2015	16.2665	21.3018	0	6	10	18	338	-0.217	0.253
	ICD-10	1954	16.2753	20.5848	0	6	10	18	262		
Denominator	ICD-9	2015	63.15	78.289	1	25	35	68	879	-0.1103	0.8592
	ICD-10	1954	64.034	78.189	1	26	36	69	852		
Rate	ICD-9	2015	0.2587	0.10865	0	0.19048	0.25	0.31818	1	-0.00193	0.4562
	ICD-10	1954	0.2555	0.1018	0	0.19048	0.25	0.3125	1		

- The developer reported that pairwise comparisons were not statistically significant for the number of numerator cases, the number of denominator cases, and the observed rates between matched hospitals between 2017 data (ICD-9) and 2018 data (ICD-10) with p-values greater than 0.05. The developer concluded this suggests that there are no differences in reliability of the measure using the previous ICD-9 coding and the current ICD-10 coding.

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.**
    - The developer provided a score-level reliability statistic indicating moderate/acceptable reliability.
    - The developer was thorough and demonstrated no difference in reliability from the switch from ICD-9 to ICD-10 codes.

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, four exclusions (multiple gestations or other presentations, not a term live birth, gestational age <37 or UTD, and previous live birth) were empirically tested for impact on the denominator. The developer provided a rationale for each exclusion and the percentage lost to the exclusions, which are not mutually exclusive. The developer further noted that, given the definition/population of interest for the measure, the specifications are incorporated into the denominator definition.
    - The developer also examined potential exclusions/codes suggested by the Society for Maternal-Fetal Medicine Health Policy and Advocacy Committee, but after empirical analyses did not add the additional codes.
    - No concerns.
    - In its previous submission, the developer noted exclusions that were not derived directly from the evidence and the justification for them.

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 289 hospitals were identified as outliers with rates beyond the two standard deviation upper limit, and 122 hospitals were identified as outliers with rates beyond the three standard deviation limits—e.g., the upper limit of a 95% confidence interval for a hospital with the median denominator size of 142 is 33.3%, and 38.1% for a 99.9% confidence interval.
    - 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

    - Hospitals transmitting data with missing data on any of the critical elements are not accepted; the measure has been collected since 2011.
    - No concerns.

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

- The developer conducted empirical analyses that indicate physician preference and subjectivity account for most of the Age and BMI effects on the nulliparous, term, singleton, vertex cesarean rate, thus supporting the lack of need for adjustment for these factors.
- The developer does not adjust for social risk factors, but does not provide a rationale for this approach (only indicated "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**

- The measure is not risk adjusted; the developer provided empirical analyses for this approach.
- The measure is not risk adjusted for social risk factors; the developer did not provide a rationale for this approach, stating only "not applicable".
- The Committee may wish to discuss these approaches with the developer.

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

- The developer used construct validity to calculate correlations of this measure with other measures of perinatal quality and with other measures of hospital quality.
    - For 0471, a lower rate is higher quality. The developer hypothesized that it would correlate negatively to perinatal care measures where a high rate is desirable (e.g., PC-05 Exclusive Breast Milk Feeding) and correlate positively to perinatal care measures where a low rate is desirable (e.g., PC-01 Elective Delivery).
    - For other measures of hospital quality, the developer used the Hospital Compare Five-Star rating system, which rates facilities on multiple quality measures. The developer hypothesized that rates for 0471 would be negatively correlated with the Five-Star rating.

- The developer stated a correlation of 0.1 - 0.3 was considered weak, 0.3 - 0.5 was considered moderate, and over 0.5 was considered strong.

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

    Submission document: Testing attachment, section 2b2.3

    - The developer reported a correlation coefficient of 0.133 with PC-01 Elective Delivery, i.e., weakly positive, but as directionally as hypothesized. A correlation of -0.28 between this measure and PC-05 Exclusive Breast Milk Feeding was reported, i.e., weakly negative but directionally as hypothesized.
    - With respect to the correlation analysis of this measure and Hospital Five-Star, the developer reported a weak negative correlation between the score for this measure (0471 Cesarean Birth) and the overall quality rating ($\rho = -0.133$, $p < .0001$). The developer noted this suggests that facilities having a higher overall quality rating tend to have a lower score on this measure, indicating higher quality in the mother's delivery.

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.**

    - No concerns.
        - The developer provided empirical data to justify its rationale not to risk adjust. Similarly, the developer provided data related to exclusions (both those it has incorporated and those it has not), and data on meaningful differences. Missing data are not an issue.
        - The developer's construct validity testing against CMS's global Five-Star system is a strong approach.

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.**
    - No additional concerns or questions.

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

2a1. Reliability – Specifications
Comments:

**In notes - "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."

**No concerns.

**No concerns with reliability

**No concerns.

**There has be a lot of controversy about whether this measure should be risk adjusted. I think transparency on the data shown in the application would help put these arguments to rest.

** I believe this measure is HIGHLY reliable, and can accept that the preliminary review states it is moderate

2a2. Reliability – Testing
Comments:

**No concerns. Mathematically I don't understand this, but the developed noted average score of 0.76, where 0.7 or higher suggests adequate reliability.

**No concerns.

**No

**Yes, given the range in the signal to noise reliability statistics, there is significant room for improvement

**none

**none

2b1. Validity –Testing
Comments:

**No

**No concerns.

**No concerns.

**Yes, unclear the utility of comparing to breast feeding and hospital compare five-star ratings

**See above in regards to case mix adjustment. For many in the obstetric community this measure trades off simplicity for face validity.

**none

2b2-3. Exclusions/Risk Adjustment
Comments:

**No concerns in the measure worksheet

**No concerns.

**Exclusions are consistent with the evidence. This measure does not risk adjust.

**no concerns

**See above re risk adjustment

**no concerns

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

Comments:

**No concerns in the measure worksheet.

**No concerns.

**Lower scores with this measure typically equate to higher quality. However, once the rate drops below a certain undetermined threshold, quality may be compromised. No concerns with missing data.

**no concerns

**N/A as missing data not accepted

**no concerns

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.

- The developer reported that data are generated by and used by healthcare personnel during the provision of care, coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), abstracted from a record by someone other than person obtaining original information. Some data elements are in defined fields in electronic sources.
- Not all hospitals currently have the capacity to abstract the electronic version of this measure, so The Joint Commission continues to offer this chart abstracted version that allows for data capture from unstructured data fields. (Specifications were updated in 2019 based on testing and current eCQM standards, and the measure will be available to hospitals in 2020 for data collection to meet accreditation requirements for eCQM submission.
- Currently, hospitals using this performance measure generally collect measure data via manual review of the EMR, data derived from vital records reports received from state or local departments of public health, delivery logs or clinical information systems or a combination. Collected data are submitted to The Joint Commission on a quarterly basis.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*

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

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

Comments:

**Some hospitals don't have this data collection capacity, but Joint Commission is try to help/support them.

**No concerns.

**For those facilities performing manual chart abstraction for this measure, the feasibility is a challenge depending on volume.

**no concerns for places that have EMRs

**no concerns

** rated as moderate, which I accept- however I believe this is HIGHLY feasible

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

- The developer reported the measure is used for accountability as part of The Joint Commission's Hospital Accreditation Program and The Joint Commission's the Perinatal Care Certification.
- The measure will be publicly reported beginning July 2020 as part of The Joint Commission's Quality Check.

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

Feedback on the measure by those being measured or others

The developer reported:

- 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.

- Most statistical questions on this measure were regarding how this measure was to be publicly reported in 2020. There was strong support for the public reporting of this measure from multiple stakeholders.
- Queries submitted via the automated feedback system have decreased significantly for this measure during the past three years.

Additional Feedback:

- Reviewed by Measure Applications Partnership in 2014; continued development recommended.

***Question for the Committee**:*

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

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

- The performance trend for this measure is as follows:
  - 2015-26.2%
  - 2016-26.1%
  - 2017-26.0%
  - 2018-25.5%
- NQF staff note that the developer reported that in January 2014, The Joint Commission required mandatory reporting of the perinatal care measure set for all accredited hospitals with 1,100 births or more annually: 1,388 hospitals reported the data with an average rate of 26.8% (n=363,400 patients). For the current submission (2018 data), 1,936 hospitals reported data (the minimum birth threshold was reduced to 300) (n=490,481 patients) and mean performance of 25.7%.

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

The developer reported two unintended consequences and the mitigating actions it took:

- Patients who did not receive prenatal care were inappropriately included in the measure denominator, as the gestational age data element was abstracted as unable to be determined (UTD).
  - Mitigating Action: To avoid penalizing hospitals, cases with UTD were removed from the measure population.
- Some hospitals have reported higher rates due to small denominator populations as a result of sampling.
  - Mitigating Action: Vital Records reports, delivery logs and clinical information systems were added as acceptable data sources to help hospitals identify all cases with =>37 weeks gestation, so that 100% of these cases could be reviewed to increase the denominator population size.

Potential harms

- None identified by the developer other than the unintended consequences above.

Additional Feedback

- None

***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 why there has been only a slight improvement between 2014 and 2018 (26.8% drop to 25.7%) and the degree to which results for the measure can drive improvement? How can the disparities data presented by the developer drive improvement?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

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

RATIONALE:

- The measure will not be publicly reported as part of Quality Check until July 2020, so unable to assess the degree to which audiences (e.g., consumers, purchasers, providers, policymakers) can or could use performance results.
- The small degree (1.1 percentage point) of improvement between 2014 and 2018 and the disparities data should be discussed re: the usability of the measure to hospitals for improvement.

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

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

Comments:

**July 2020 public reporting will start.

**This measure will begin to be publicly reported in July 2020. Those being measured have had opportunity to provide feedback and to learn how to report on this measure.

**Performance with this measure is being used by some private payors for tiering and other accountability programs. Those being measured are able to offer feedback on the measure.

**no concerns, appears to be a priority among stakeholders

**Data is given on the ability to combine this measures with others to give an overall rating of perinatal quality.

** need mandatory reporting- which is in process.

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

Comments:

**All providers who provide labor support are always worried about the potential for neonatal death - although the data presented at the beginning of the measure worksheet are compelling and reassuring (no worse outcomes). I am interested to hear why the developers thought more progress wasn't made and how they think the public reporting will change things.

**This measure is especially important in reducing severe maternal morbidity and mortality.

**Harm may occur when rates are too low.

**Any potential harms seem to have been addressed, and to increase the benefit, reporting disparities data seems important

**none

** highly usable, data supports IMPROVEMENTS even in balancing measures (NICU admissions, LOS, patient satisfaction)

Criterion 5: Related and Competing Measures

Related or competing measures

- None

Harmonization

- Not applicable

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

Comments:

**no

**No concerns.

**None.

**n/a

**consider bundling with a measure of neonatal outcome

** well harmonized AND, important because it addresses overuse, misuse, underuse issues

Public and Member Comments

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

- No comments received
- **Of the 0 NQF members who have submitted a support/non-support choice:**
  - 0 support the measure
  - 0 do not support the measure

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_PC02_0471.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*)**:** 0471

Measure Title: PC-02 Cesarean Birth

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: Nulliparous women with a term, singleton baby in a vertex position delivered by cesarean birth

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

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

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

☐ Process: Click here to name what is being measured

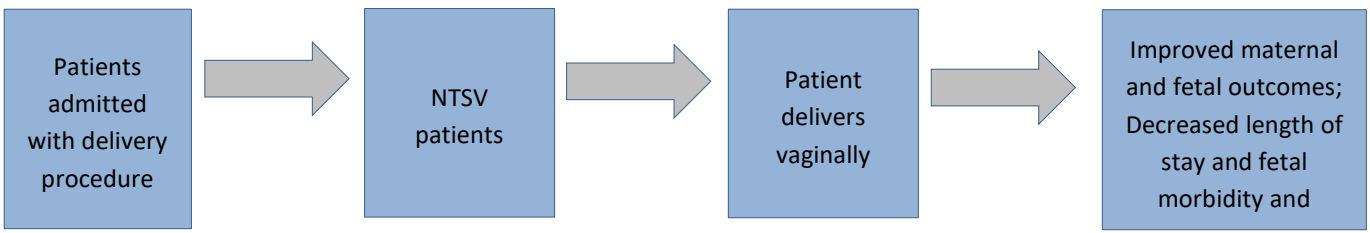
☐ Appropriate use measure: Click here to name what is being measured

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

1a.2 LOGIC MODEL Outcome Complete for all measures

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 cesarean deliveries >> population determined in nulliparous, term singleton newborns in vertex position>> population assessed; nulliparous term patients>> patient delivers vaginally >> 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.)

****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.

Safe Prevention of the Primary Cesarean Delivery
Caughey et al., 2019

The intent of this measure is to reduce the rates of cesarean sections, and as a result, reduce the number of complications and negative outcomes for moms and babies. Caughey et al., 2019, found this reduction is consistent with that of the recommendations given by the American College of Obstetricians and Gynecologists. In 2011, one in three women who gave birth in the United States did so by cesarean delivery. Even though the rates of primary and total cesarean delivery have plateaued recently, there was a rapid increase in cesarean rates from 1996 to 2011. Although cesarean delivery can be lifesaving for the fetus, the mother, or both in certain cases, the rapid increase in the rate of cesarean births without evidence of concomitant decreases in maternal or neonatal morbidity or mortality raises significant concern that cesarean delivery is overused. Therefore, it is important for health care providers to understand the short-term and long-term tradeoffs between cesarean and vaginal delivery, as well as the safe and appropriate opportunities to prevent overuse of cesarean delivery, particularly primary cesarean delivery.

The recommendations by the American College of Obstetricians and Gynecologists have identified 40 studies that help to support the evidence in reducing the number of nulliparous, term, singleton newborns in vertex position to be delivered by cesarean birth. Of those studies, there were Population-based, Retrospective-Cohort, Prospective, Randomized trials, and Meta-analysis studies performed. Some of those recommendations include: allowing for increased length of time for pushing during the active stage of labor (at least 2 hours of pushing in multiparous women and 3 hours of pushing for nulliparous patients) and operative vaginal delivery in the second stage of labor by experienced and well-trained physicians should be considered a safe, acceptable alternative to cesarean delivery. Training in, and ongoing maintenance of, practical skills related to operative vaginal delivery should be encouraged (forceps or manual rotation/aversion). Several approaches are needed to reduce the primary cesarean delivery rate, which in turn would lower the repeat cesarean delivery rate. Although national and regional organizations can take the lead in setting the agenda regarding the safe prevention of primary cesarean delivery, such an agenda will need to be prioritized at the level of practices, hospitals, health care systems, and, of course, patients.

A 2007 review found that the cesarean delivery rate was reduced by 13% when audit and feedback were used exclusively and decreased by 27% when audit and feedback were used as part of a multifaceted intervention, which involved second opinions and culture change. Systemic interventions, therefore, provide an important strategic opportunity for reducing cesarean delivery rates. However, the specific interventional approaches have not been studied in large, prospective trials, thus specific recommendations cannot be made.

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

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

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

Safety Assessment Scale of a Large-Scale Improvement Collaborative to Reduce Nulliparous Cesarean Delivery Rates.
Main, E.K et al., 2019

In 2015, nulliparous, term, singleton, vertex cesarean delivery rate ranged from 11.3% to 76.9% in 248 California hospitals with maternity services. Fifty-six hospitals, all among those with initial rates above 23.9% participated in CMQCC's Supporting Vaginal Birth collaborative. The overall nulliparous, term, singleton, vertex cesarean delivery rates declined by 4.5 percentage points (a relative decline of 15.5%), from 29.1% in the first two quarters in 2015 to 24.6% in the last two quarters in 2017. The largest decline happened in the third and fourth quarters of 2016, when the collaborative initiated. Twenty-four of the 56 participating hospitals lowered their nulliparous, term, singleton, vertex cesarean delivery rates to below the Healthy People 2020 target of 23.9% in 2017. Ten of them lowered the rates even further, to below 20% (range 15.0–19.9%). Similar to the analysis based on the absolute amount of decline shown above, we did not observe an increase in adverse maternal or neonatal outcomes even among those with these lower final rates (hospitals with 2017 nulliparous, term, singleton, vertex cesarean delivery rate between 15.0% and 19.9%). Rates of severe unexpected newborn complications also declined from 2.5% in 2015 to 2.2% in 2017 in this group, but not significantly with an adjusted OR of 0.84 (95% CI 0.58–1.20). The size of this study is noteworthy. Fifty-six hospitals participated with a total annual delivery volume of 119,000 women. This is a higher delivery volume than in all but nine U.S. states. All hospitals in the collaborative had a starting nulliparous, term, singleton, vertex cesarean delivery rate higher than the Healthy People 2020 national target of 23.9%.18 Importantly, the majority of collaborative hospitals were community hospitals (87%), representing the predominant care model in the United States. The CMQCC Supporting Vaginal Birth collaborative interventions emphasized reducing

latent phase cesarean deliveries, implementation of ACOG–SMFM guidelines for diagnosis and management of active phase disorders, and enhanced nursing support (increased walking and upright positioning, use of peanut balls, and interpersonal coaching). Providers did not increase their operative vaginal delivery rates. Findings provide evidence that a reduction in first birth cesarean delivery rates need not be associated with more difficult vaginal births or higher rates of major perineal lacerations. The large number of nulliparous, term, singleton, vertex deliveries provided the ability to confidently examine maternal and neonatal complications that are infrequent. The range of improvement among the 56 hospitals created the opportunity to perform a sensitivity analysis comparing hospitals with very high levels of cesarean delivery rate reduction (217.1 to 27.1 percentage points) with those with limited change (22.4 to +4.7 percentage points). The rate of severe unexpected newborn complications (the major composite index for neonatal outcomes) actually improved in hospitals with the greatest reduction in nulliparous, term, singleton, vertex cesarean delivery rate. This is in concordance with several of the single hospital studies noted.

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

Not applicable

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)

Not applicable

☐Clinical Practice Guideline recommendation

☐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

2016 submission

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	American College of Obstetricians and Gynecologists (ACOG) Task Force on Cesarean Delivery Rates. Evaluation of Cesarean Delivery. 2000. Washington, D.C. The American College of Obstetricians and Gynecologists (ACOG) is the nation´s leading group of professionals providing health care for women. The monograph developed by the ACOG Task Force on Cesarean Delivery provides 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. Specifically, NTSV cesarean section rate is preferred over total or primary cesarean rates as it more narrowly focuses on the population at greatest risk (nulliparous women in labor) with the greatest long-term consequences.
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.	The American College of Obstetricians and Gynecologists (ACOG) in their monograph on evaluating Cesarean delivery rates, recommended this measure for benchmarking cesarean section rates on page 35: "Institutions and practitioners should consider reviewing their cesarean delivery rates with these benchmarks for 1) nulliparous women with term singleton fetuses with vertex presentations and, 2) multiparous women with one previous low-transverse cesarean delivery and term singleton fetuses with vertex presentations." II-3
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 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	
Grade assigned to the **recommendation** with definition of the grade	Yes, grading was assigned to the recommendations. The American College of Obstetrics and Gynecology 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 in the number of nulliparous patients with live term singleton newborns in vertex position delivering by cesarean section. The target population for the performance measure is consistent with the body of evidence supporting quality improvement strategies to reduce the number of NTSV cesarean sections. **Quantity:** The body of literature looking at cesarean section rates is very large with over 5,000 articles published since 1980. Over 1,000 articles have focused on the quality issues around nulliparous (sometimes called primiparous) cesarean section rates. Specifically, the low-risk first-birth cesarean rate has been examined by over 250 retrospective cohort and prospective observational studies. Synonyms in the literature include: NTSV (nulliparous, term, singleton, vertex) cesarean rate, Standard primip cesarean rate, and the Robson 10-category cesarean classification system (of which NTSV is the key driver). Throughout these studies, NTSV cesarean has emerged as the group with highest

	variation and greatest contribution to the rise in cesarean rates both in the US and internationally. **Quality:** The quality of evidence supporting the reduction in the number of NTSV cesarean sections is quite high. The rate of severe obstetric hemorrhage has significantly increased (by 50%) over the last 15 years in the U.S. There has also been a 270% increase in blood transfusions, with both hemorrhage and transfusions correlated to the rise in cesarean deliveries. Infection is the most common serious complication of cesarean delivery with typical rates of 3 to 9%. As noted, ACOG has evaluated cesarean sections and made a recommendation to adopt the NTSV cesarean section rate as a national metric to address through quality improvement interventions. Studies of quality improvement initiatives aimed at reducing NTSV cesarean sections have also noted a decrease in the number of such deliveries as well as a subsequent decrease in the number of maternal and neonatal morbidities. There is no documented evidence regarding controversy related to the reduction of NTSV cesarean sections. A review of recent studies also supports the use of quality improvement interventions to further reduce the number of such deliveries. **Quantity:** High **Quality:** High **Consistency:** High
Estimates of benefit and consistency across studies	The body of evidence consistently supports the reduction of NTSV cesarean deliveries. Studies looking at multi-faceted quality improvement interventions also show a decrease in the number of NTSV cesarean sections. NTSV cesarean delivery rates show much more consistency than total or primary cesarean delivery rates as it is much more tightly focused on labor management issues. As described before, nulliparous patients with live term singleton newborns in vertex position delivering vaginally result in improved maternal and neonatal outcomes and will result in substantial savings in health care costs. Furthermore, the benefit is extended to all future pregnancies—if the first birth is a cesarean, then 90% of the remainder will be cesareans with health risks markedly increasing for each additional cesarean.
What harms were identified?	
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	

1a.4 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please

describe the evidence on which you are basing the performance measure.

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

2020 submission: Citations from Evidence Other Than Guidelines

- Caughey, A.B., Cahill, A.G., Guise, JM., Rouse, D.J. (2019). Safe prevention of the primary cesarean delivery. *American College of Obstetricians and Gynecologists,* 123: 693-711. Retrieved from https://www.acog.org/Clinical-Guidance-and-Publications/Obstetric-Care-Consensus-Series/Safe-Prevention-of-the-Primary-Cesarean-Delivery.
- Main, E.K, Shen-Chih, C., Cape, V., Sakowski, C., Smith, H., Vasher, J. (2019). Safety assessment scale of a large-scale improvement collaborative to reduce nulliparous cesarean delivery rates. *American College of Obstetricians and Gynecologists,* 133 (4): 613-623.

From previous submission: Citations from Evidence Other Than Guidelines

• Barber EL, Lundsberg LS, Belanger K, Pettker CM, Funai EF, Illuzzi JL. (2011). Indications contributing to the increasing cesarean delivery rate. Am J Obstet Gynecol. 118(1):29-38.
• Brennan, DJ, Robson, MS, Murphy, M, O´Herlihy, C. (2009). Comparative analysis of international cesarean delivery rates using 10-group classification identifies significant variation in spontaneous labor. American Journal of Obstetrics and Gynecology. 201(3):308 e301-308.
• Coonrod, DV, Drachman, D, Hobson, P, Manriquez, M. (2008). Nulliparous term singleton vertex cesarean delivery rates: institutional and individual level predictors. American Journal of Obstetrics and Gynecology. 198(6):694 e691-611; discussion 694 e611.
• Ehrenthal, DB, Jiang, X, & Strobino, DM. (2010). Labor induction and the risk of a cesarean delivery among nulliparous women at term. Am J Obstet Gynecol . 116(1):35-42.
• Getahun D, Strickland D, Lawrence JM, Fassett MJ, Koebnick C, Jacobsen SJ. (2009). Racial and ethnic disparities in the trends in primary cesarean delivery based on indications. Am J Obstet Gynecol. 201(4):422 e421-427.
• Main, EK, Moore, D, Farrell, B, et al. (2006). Is there a useful cesarean birth measure? Assessment of the nulliparous term singleton vertex cesarean birth rate as a tool for obstetric quality improvement. American Journal of Obstetrics and Gynecology. 194(6):1644-1651; discussion 1651-1642.
• Main, EK. (1999). Reducing cesarean birth rates with data-driven quality improvement activities. Pediatrics.103(1 Suppl E):374-383.
• US Department of Health and Human Services (DHHS). (2000). Healthy People 2010. Washington, DC. Retrieved on September 26, 2011 at: http://www.healthypeople.gov/2010

1b. Performance Gap

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

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

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

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

The removal of any pressure to not perform a cesarean birth has led to a skyrocketing of hospital, state and national cesarean birth (CB) rates. Some hospitals now have CB rates over 50%. Hospitals with CB rates at 15-20% have infant outcomes that are just as good and better maternal outcomes (Gould et al., 2004). There are no data that higher rates improve any outcomes, yet the CB rates continue to rise. This measure seeks to focus attention on the most variable portion of the CB epidemic, the term labor CB in nulliparous women. This population segment accounts for the large majority of the variable portion of the CB rate, and is the area most affected by subjectivity.

As compared to other CB measures, what is different about NTSV CB rate (Low-risk Primary CB in first births) is that there are clear cut quality improvement activities that can be done to address the differences. Main et al. (2006) found that over 60% of the variation among hospitals can be attributed to first birth labor induction rates and first birth early labor admission rates. The results showed if labor was forced when the cervix was not ready the outcomes were poorer. Alfirevic et al. (2004) also showed that labor and delivery guidelines can make a difference in labor outcomes. Many authors have shown that physician factors, rather than patient characteristics or obstetric diagnoses are the major driver for the difference in rates within a hospital (Berkowitz, et al., 1989; Goyert et al., 1989; Luthy et al., 2003). The dramatic variation in NTSV rates seen in all populations studied is striking according to Menacker (2006). Hospitals within a state (Coonrod et al., 2008; California Office of Statewide Hospital Planning and Development [OSHPD], 2007) and physicians within a hospital (Main, 1999) have rates with a 3-5 fold variation.

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

The measure will assist health care organizations (HCOs) to track nulliparous patients with live term singleton newborns in vertex position delivering by cesarean birth to reduce the occurrence. Nulliparous women have 4-6 times the cesarean birth rate than multiparous women thus the NTSV population is the largest driver of primary cesarean birth rate. Furthermore, nulliparity varies greatly among hospitals (20% to 60%) making it the most important risk factor for stratification or adjustment. NTSV has the large variation among facilities thus identifying an important population on which to focus quality improvement efforts.

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

Sources

- Agency for Healthcare Research and Quality. (2002). AHRQ Quality Indicators Guide to Inpatient Quality Indicators: Quality of Care in Hospitals Volume, Mortality, and Utilization. Revision 4 (December 22, 2004). AHRQ Pub. No. 02-RO204.
- Alfirevic, Z., Edwards, G., & Platt, M.J. (2004). The impact of delivery suite guidelines on intrapartum care in “standard primigravida.” Eur J Obstet Gynecol Reprod Biol.115:28-31.
- American College of Obstetricians and Gynecologists. (2000). Task Force on Cesarean Delivery Rates. Evaluation of Cesarean Delivery. (Developed under the direction of the Task Force on Cesarean Delivery Rates, Roger K. Freeman, MD, Chair, Arnold W. Cohen, MD, Richard Depp III, MD, Fredric D. Frigoletto Jr, MD, Gary

D.V. Hankins, MD, Ellice Lieberman, MD, DrPH, M. Kathryn Menard, MD, David A. Nagey, MD, Carol W. Saffold, MD, Lisa Sams, RNC, MSN and ACOG Staff: Stanley Zinberg, MD, MS, Debra A. Hawks, MPH, and Elizabeth Steele).

- Bailit, J.L., Garrett, J.M., Miller, W.C., McMahon, M.J., & Cefalo, R.C. (2002). Hospital primary cesarean delivery rates and the risk of poor neonatal outcomes. Am J Obstet Gynecol. 187(3):721-7.
- Bailit, J. & Garrett, J. (2003). Comparison of risk-adjustment methodologies. Am J Obstet Gynecol.102:45-51.
- Bailit, J.L., Love, T.E., & Dawson, N.V. (2006). Quality of obstetric care and risk-adjusted primary cesarean delivery rates. Am J Obstet Gynecol.194:402.
- Bailit, J.L. (2007). Measuring the quality of inpatient obstetrical care. Ob Gyn Sur. 62:207-213.
- Berkowitz, G.S., Fiarman, G.S., Mojica, M.A., et al. (1989). Effect of physician characteristics on the cesarean birth rate. Am J Obstet Gynecol. 161:146-9.
- California Office of Statewide Hospital Planning and Development. (2017). Hospital Volume and Utilization Indicators for California, Retrieved from the Internet on February 22, 2018 at: https://www.oshpd.ca.gov/HID/AHRQ-Volume-Utilization.html
- Cleary, R., Beard, R.W., Chapple, J., Coles, J., Griffin, M., & Joffe, M. (1996). The standard primipara as a basis for inter-unit comparisons of maternity care. Br J Obstet Gynecol. 103:223-9.
- Coonrod, D.V., Drachman, D., Hobson, P., & Manriquez, M. (2008). Nulliparous term singleton vertex cesarean delivery rates: institutional and individual level predictors. Am J Obstet Gynecol. 694-696.
- DiGiuseppe, D.L., Aron, D.C., Payne, S.M., Snow, R.J., Dieker, L., & Rosenthal, G.E. (2001). Risk adjusting cesarean delivery rates: a comparison of hospital profiles based on medical record and birth certificate data. Health Serv Res.36:959-77.
- Gould, J., Danielson, B., Korst, L., Phibbs, R., Chance, K.,& Main, E.K., et al. (2004). Cesarean delivery rate and neonatal morbidity in a low-risk population. Am J Obstet Gynecol, 104:11-19.
- Goyert, G.L., Bottoms, F.S., Treadwell, M.C., et al. (1989). The physician factor in cesarean birth rates. N Engl J Med.320:706-9.
- Le Ray, C., Carayol, M., Zeitlin, J., Berat, G., & Goffinet, F. (2006). Level of perinatal care of the maternity unit and rate of cesarean in low-risk nulliparas. Am J Obstet Gynecol. 107:1269-77.
- Luthy, D.A., Malmgren, J.A., Zingheim, R.W., & Leininger, C.J. (2003). Physician contribution to a cesarean delivery risk model. Am J Obstet Gynecol.188:1579-85.
- Main, E.K. (1999). Reducing cesarean birth rates with data-driven quality improvement activities. Peds. 103: 374-383.
- Main E.K., Bloomfield, L., & Hunt, G. (2004). Development of a large-scale obstetric quality-improvement program that focused on the nulliparous patient at term. Am J Obstet Gynecol.190:1747-58.
- Main, E.K., Moore, D., Farrell, B., Schimmel, L.D., Altman, R.J., Abrahams, C., et al., (2006). Is there a useful cesarean birth measure? Assessment of the nulliparous term singleton vertex cesarean birth rate as a tool for obstetric quality improvement. Am J Obstet Gynecol. 194:1644-51.
- Menacker, F. (2005).Trends in cesarean rates for first births and repeat cesarean rates for low-risk women: United States, 1990-2003. Nat Vital Stat Rep. 54(4): 1-5.
- Romano, P.S., Yasmeen, S., Schembri, M.E., Keyzer, J.M., & Gilbert, W.M. (2005). Coding of perineal lacerations and other complications of obstetric care in hospital discharge data. Am J Obstet Gynecol.106:717-25.
- U.S. Department of Health and Human Services. (2000). Healthy People 2010: Understanding and Improving Health. 2nd ed. Washington, DC: U.S. Government Printing Office. Measure 16-9.

• Yasmeen, S., Romano, P.S., Schembri, M.E., Keyzer, J.M., & Gilbert, W.M. (2006). Accuracy of obstetric diagnoses and procedures in hospital discharge data. Am J Obstet Gynecol. 194:992-1001.

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.*

2020 Submission

CY 2018: Recent data continues to show considerable variability, with over half of hospitals reporting rates above The Healthy People 2020 goal of 23.9%. Statistics for 2018 discharges are as follows:

CY 2018 Statistics:

Number of hospitals: 1936

Total Number of Patients: 490,481

Mean (SD): 25.7% (7.6%)

IQR: 8.9%

Deciles (0,10,20,30,40,50,60,70,80,90,100): 2.0%, 16.9%, 19.7%, 21.8%, 23.5%, 25.0%, 26.8%, 28.6%, 30.9%, 34.8%, 100%

Correction

The trend for this measure shows a significant decrease in the rate in 2018. Recent data continues to show considerable variability, with over half of hospitals reporting rates above the Healthy People 2020 goal of 23.9%.

2015-26.2%

2016-26.1%

2017-26.0%

2018-25.5%

2016 Submission

Nulliparous term singleton vertex cesarean births continue to remain above The Healthy People 2020 goal of 23.9% (DHHS, 2010). The Perinatal Care (PC) core measures were added as a new core measure set in 2010 for hospitals to select in order to meet their ORYX performance measurement requirement for Joint Commission accreditation purposes. At that time, approximately 165 hospitals reported the data with an average measure rate of 26.7% (n=25,143 patients). In January 2014, The Joint Commission required mandatory reporting of the PC measure set for all accredited hospitals with 1100 births or more annually. 1388 hospitals reported the data with an average rate of 26.8% (n=363,400 patients). The 2014 performance gap persists with improvement noted primarily in the lower quartile (21.1%) and 10th percentile (17.6%) hospitals. It is important to note that a performance gap of 12.4% exists for the 90th percentile of hospitals performing at 36.3% (if 23.9% is considered goal performance). The 2014 mean rate of 26.7% also remains above the HP 2020 goal. The threshold for mandatory reporting was recently lowered to 300 births annually effective January 2016. The new reporting requirement will now capture approximately 80% of all accredited birthing hospitals. As a result, the rates increased with the addition of approximately 821 more hospitals reporting data. Below is the specified level of analysis for PC-02 beginning with discharges April 1, 2010 through December 31, 2014.

• 2Q 2010: 25,143 denominator cases; 6,708 numerator cases; 165 hospitals; 26.7% national aggregate rate; 0.26636 mean of hospital rates; 0.09659 standard deviation; 40.0% 90th percentile rate; 31.9% 75th percentile rate/upper quartile; 26.3% 50th percentile rate/median rate; 20.5% 25th percentile rate/lower quartile; and 15.4% 10th percentile rate.

- CY 2011: 33,379 denominator cases; 8,779 numerator cases; 166 hospitals; 26.3% national aggregate rate; 0.26283 mean of hospital rates; 0.08961 standard deviation; 35.3% 90th percentile rate; 30.7% 75th percentile rate/upper quartile; 25.7% 50th percentile rate/median rate; 20.8% 25th percentile rate/lower quartile; and 16.8% 10th percentile rate.
- CY 2012: 33,944 denominator cases; 9,428 numerator cases; 169 hospitals; 26.2% national aggregate rate; 0.26335 mean of hospital rates; 0.08582 standard deviation; 36.1% 90th percentile rate; 31.1% 75th percentile rate/upper quartile; 25% 50th percentile rate/median rate; 20.2% 25th percentile rate/lower quartile; and 17.1% 10th percentile rate.
- CY 2013: 44,679 denominator cases; 11,553 numerator cases; 200 hospitals; 25.9% national aggregate rate; 0.25792 mean of hospital rates; 0.09181 standard deviation; 35.7% 90th percentile rate; 30.8%% 75th percentile rate/upper quartile; 25% 50th percentile rate/median rate; 20% 25th percentile rate/lower quartile; and 16.4% 10th percentile rate.
- CY 2014: 363,400 denominator cases; 97,270 numerator cases; 1388 hospitals; 26.8% national aggregate rate; 0.26732 mean of hospital rates; 0.09064 standard deviation; 36.3% 90th percentile rate; 31.0% 75th percentile rate/upper quartile; 25.9% 50th percentile rate/median rate; 21.2% 25th percentile rate/lower quartile; and 17.6% 10th percentile rate.

Note: PC-02 hospital rates listed in this section were not age standardized.

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.*

For 2018 discharges:

Rates by Age category

Age	Rate (%)
<20	16.1
20-24	21.6
25-29	25.2
30-34	28.9
35-39	38.1
40+	53.0

Rates by Hispanic Ethnicity

Hispanic Ethnicity	Rate (%)
No	24.9
Yes	24.7

Rates by Race

Race	Rate (%)
White	24.1
African American	29.0
American Indian	24.3
Asian	25.0
Pacific Islander	26.7
Unable to Determine	24.2

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

Not applicable

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):*

Disparities Sensitive, 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://manual.jointcommission.org/releases/TJC2020A2/

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 not an eMeasure **Attachment:**

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:** PC02AppendixATJCTablesv2020A2.xlsx

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

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

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

Not an instrument-based measure

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

Yes

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

Updated data element Gestational Age: Notes for abstraction and Suggested Data Sources have been updated and reordered to clarify, reduce burden of abstraction and align with the eCQM measure specifications.

Updated data element Prior Uterine Surgery: Added notes in order to clarify abstraction of prior uterine surgeries.

Appendix A - ICD-10 Code Tables: Revised to reflect the ICD-10 code updates for Fiscal Year (FY) 2019, effective for discharges October 1, 2018

Updated data elements: Data element Number of Previous Live Births replaced with the new data element Previous Live Births to allow for capture of nulliparous by a yes or no allowable value and to reduce the burden of abstracting the actual number of previous live births.

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.*

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

Patients with cesarean births with ICD-10-PCS Principal Procedure Code or ICD-10-PCS Other Procedure Codes for cesarean birth as defined in Appendix A, Table 11.06.

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)*

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

Two data elements are used for the observed outcome and to calculate the numerator:

1. ICD-10-PCS Other Procedure Codes - The International Classification of Diseases, Tenth Revision, Procedure Coding System code that identifies significant procedures performed other than the principal procedure during this hospitalization.

2. ICD-10-PCS Principal Procedure Code - The International Classification of Diseases, Tenth Revision, Procedure Coding System (ICD-10-PCS) code that identifies the principal procedure performed during this hospitalization. The principal procedure is the procedure performed for definitive treatment rather than diagnostic or exploratory purposes, or which is necessary to take care of a complication.

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

The outcome target population being measured is: Nulliparous patients with an ICD-10-CM Principal or Other Diagnosis Code for outcome of delivery as defined in Appendix A, Table 11.08 and with a delivery of a newborn with 37 weeks or more gestation completed or with an ICD-10-PCS Principal or Other Procedure Codes for delivery as defined in Appendix A, Tables 11.01.1.

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).

Seven data elements are used to identify the outcome target population and to calculate the denominator:

1. Admission Date – The month, day, and year of admission to acute inpatient care.

2. Birthdate - The month, day, and year the patient was born.

3. Discharge Date – The month, day, and year the patient was discharged from acute care, left against medical advice, or expired during the stay.

4. Gestational Age – Documentation of the weeks of gestation completed at the time of delivery. Allowable Values: 1-50 or UTD.

5. ICD-10-CM Other Diagnosis Codes - The International Classification of Diseases, Tenth Revision, Clinical Modification codes associated with the other or secondary diagnoses for this hospitalization.

6. ICD-10-CM Principal Diagnosis Code - The International Classification of Diseases, Tenth Revision, Clinical Modification diagnosis code that is primarily responsible for the admission of the patient to the hospital for care during this hospitalization.

7. Number of Previous Live Births - The number of deliveries resulting in a live birth the patient experienced prior to current hospitalization. Allowable Values: 0-50 or UTD (as of 1/1/2019 Previous Live Births - Documentation that the patient experienced a live birth prior to the current hospitalization. Allowable values: Yes or No/UTD.)

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

• ICD-10-CM Principal Diagnosis Code or ICD-10-CM Other Diagnosis Codes for multiple gestations and other presentations as defined in Appendix A, Table 11.09

• Less than 8 years of age

• Greater than or equal to 65 years of age

• Length of Stay >120 days

• Gestational Age < 37 weeks or UTD

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.)*

• Patients with ICD-10-CM Principal Diagnosis Code or Other Diagnosis Codes for multiple gestations and other presentations are excluded. Appendix A, Table 11.09

• The patient age in years is equal to the Admission Date minus the Birthdate. Patients less than 8 years of age or greater or equal to 65 years of age are excluded.

• Length of stay (LOS) in days is equal to the Discharge Date minus the Admission Date. If the LOS is greater than 120 days, the patient is excluded.

• Patients with a Gestational Age less than 37 weeks or UTD are excluded from the measure.

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

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.)*

1. Start processing. Run cases that are included in the PC-Mother Initial Patient Population and pass the edits defined in the Transmission Data Processing Flow: Clinical through this measure.

2. Check ICD-10-CM Principal or Other Diagnosis Codes

a) If at least one of the ICD-10-CM Principal or Other Diagnosis Code is on Table 11.09, the case will proceed to a Measure Category Assignment of B and will not be in the measure population. Stop processing.

b) If none of the ICD-10-CM Principal or Other Diagnosis Code is on Table 11.09, continue processing and proceed to recheck ICD-10-CM Principal or Other Diagnosis Codes.

3. Recheck ICD-10-CM Principal or Other Diagnosis Codes

a) If none of the ICD-10-CM Principal or Other Diagnosis Codes is on Table 11.08, the case will proceed to a Measure Category Assignment of B and will not be in the measure population. Stop processing.

b) If at least one of the ICD-10-CM Principal or Other Diagnosis Codes is on Table 11.08, continue processing and proceed to Gestational Age.

4. Check Gestational Age

a) If Gestational Age is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.

b) If Gestational Age is less than 37 or equal to an Unable to Determine Value, the case will proceed to a Measure Category Assignment of B and will not be in the measure population. Stop processing.

c) If Gestational Age is greater than or equal to 37, continue processing and proceed to Number of Previous Live Births.

5. Check Previous Live Births

a) If Previous Live Births is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop Processing.

b) If Previous Live Births is Yes, the case will proceed to a Measure Category Assignment of B and will not be in the measure population. Stop Processing.

c) If Previous Live Births is No, continue processing and proceed to recheck ICD-10- CM Principal Procedure or Other Diagnosis Codes.

6. Check ICD-10-PCS Principal or Other Procedure Codes

a) If all of the ICD-10-PCS Principal or Other Procedure Codes are missing or none of the ICD-10-PCS Principal or Other Procedure Codes is on Table 11.06, the case will proceed to a Measure Category Assignment of D and will be in the measure population. Stop processing.

b) If at least one of the ICD-10-PCS Principal or Other Procedure Code is on Table 11.06, the case will proceed to a Measure Category Assignment of E and will be in the Numerator Population. Stop processing.

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.

The initial patient population includes patients admitted to the hospital for inpatient acute care for deliveries. Patients are included if they have: ICD-10-PCS Principal or Other Procedure Code as defined in Appendix A, Table 11.01.1, a Patient Age (Admission Date – Birthdate) >= 8 years and < 65 and a Length of Stay (Discharge Date - Admission Date) = 120 days. The sample is taken randomly as follows for a monthly sample:

• Average monthly Initial Patient Population >= 501 results in a minimum random sample size of 101.

• Average monthly Initial Patient Population 126 – 500 results in a minimum random sample size of 20% of the population size.

• Average monthly Initial Patient Population 25 – 125 results in a minimum random sample size of 25.

• Average monthly Initial Patient Population < 25 results in no sampling; 100% Initial Patient Population required

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

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

If other, please describe in S.18.

Electronic Health Records, Other, Paper Medical Records

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

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

Starting in 2020, hospitals will use the Direct Data Submission Platform for submission of chart abstracted measures. Thus, in 2020, organizations have one place to submit both eCQM and chart abstracted data. The goal of the Direct Data Submission Platform is to ease the burden and expense of submission and empower organizations with data for quality improvement.

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

2020_nqf_testing_attachment_PC02_0471_final-637227328605591311.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*)**:** 0471

Measure Title: PC-02 Cesarean Birth

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
This submission included initial testing of measure score reliability and validity was performed using data from hospital discharges occurring in 1Q2011. Also, updates to validity and exclusions used 2015 data.

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, healthplan*)

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 nulliparous women with a term, singleton baby in a vertex position delivered by cesarean birth. 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 1936 hospitals submitting the measure using 2018 discharges and had greater than or equal to 30 denominator cases, the minimum sample size required for public reporting. The hospitals were geographically diverse and varied in size.

1936 health care organizations representing various types, locations and sizes:
362 For Profit, 1316 Not for Profit, 258 Government
627>=300 beds; 900 100-300 beds; 409 <100 beds
403 Rural; 1533 Urban

193 Major Teaching; 919 Minor Teaching; 824 Non-Teaching

2016 Submission
The PC measure set has been in national use since the 2nd quarter of 2010. It is a requirement of participation in the ORYX initiative that data on all measures in the set are collected. (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.)
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

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
Testing data: 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.

Median denominator size for the Cesarean birth measure, 2018 deliveries (Number of patients=14,184)

Number of Hospitals	Median number of deliveries	Median number of denominator cases
1936	1091	142

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 reliability and validity 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 element; 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
Reliability testing of performance measure score
We utilized the Beta-binomial model (Adams 2009) to assess how well one can distinguish the performance of one hospital from another. Conceptually, the Beta-binomial model measures the ratio of signal to noise. The signal is the proportion of the variability in the measured performance that can be explained by real differences in performance. The Beta-binomial model is an appropriate model when estimating the reliability of simple pass/fail rate measures as is the case with most Joint Commission measures. Reliability scores range from 0.0 to 1.0. A score of zero implies that all variation is attributed to measurement error (i.e., noise), whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance (across hospitals).

Adams, J.L. The Reliability of Provider Profiling: A Tutorial. Santa Monica, California: RAND Corporation. TR-653-NCQA, 2009

Comparison of ICD-9 to ICD-10 codes
Reliability was measured by comparing results from the 2017 data submission that utilized ICD-9 codes to the results from the 2018 data submission that utilized ICD-10 codes. Summary statistics for the number of numerator cases, number of denominator cases, and observed rates are presented. Hospital data were also matched by each year in each of the three attributes and a paired t-test was used to determine statistical significance for each attribute.

This measure was adapted from NQF-endorsed measure 0471 Cesarean Rate for Low-Risk First Birth Women (NTSV CS Rate). As such, initial data reliability would have been addressed during the original endorsement. The Joint Commission will be conducting additional reliability studies on this measure as well as the entire PC measure set beginning in October 2011.

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 re-abstracted with originally abstracted data having been blinded so that the re-abstraction is not biased.
- Re-abstracted 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*)

2020 Submission

Reliability testing of performance measure score
Reliability statistic for the measure: proportion of nulliparous women with a term, singleton baby in a vertex position delivered by cesarean birth:
Average: 0.76
Median: 0.76
10^{th}-90^{th} percentile across hospitals: 0.63 – 0.92

Comparison of ICD-9 to ICD-10 codes

Comparison between the ICD-9 and ICD-10 number of numerator cases, denominator cases, and observed rates

		N	Mean	Std. Dev.	Min	Q1	Median	Q3	Max	Pairwise Difference	P-Value
Numerator	ICD-9	2015	16.2665	21.3018	0	6	10	18	338	-0.217	0.253
	ICD-10	1954	16.2753	20.5848	0	6	10	18	262		
Denominator	ICD-9	2015	63.15	78.289	1	25	35	68	879	-0.1103	0.8592
	ICD-10	1954	64.034	78.189	1	26	36	69	852		
Rate	ICD-9	2015	0.2587	0.10865	0	0.19048	0.25	0.31818	1	-0.00193	0.4562
	ICD-10	1954	0.2555	0.1018	0	0.19048	0.25	0.3125	1		

Data element agreement rates 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-02.

Data Elements with a Mismatch	Total Numerator	Total Denominator	Rate
Gestational Age	639	712	89.75%
Parity	492	505	97.43%

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?*)

2020 Submission

In general, a score of 0.7 or higher suggests the measure has adequate reliability. The results suggest the measure has acceptable reliability for most of the hospitals.

Pairwise comparisons were not statistically significant for the number of numerator cases, the number of

denominator cases, and the observed rates between matched hospitals between 2017 data (ICD-9) and 2018 data (ICD-10) with p-values greater than 0.05. This suggests that there are no differences in reliability of the measure using the previous ICD-9 coding and the current ICD-10 coding.

2b1. VALIDITY TESTING-New

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

Correlations of the measure with other measures of perinatal care quality. Since a low measure rate for PC-02 is desirable, this measure is hypothesized to correlate negatively to other perinatal care measures where a high rate is desirable (PC-05, ePC-05) and correlate positively to perinatal care measures where a low rate is desirable (PC-01, ePC-01).

A correlation of 0.1 - 0.3 was considered weak, 0.3 - 0.5 was considered moderate, and over 0.5 was considered strong.

Correlation with other measures of hospital facility quality. Hospital Compare uses a five-star rating system to rate facilities based on multiple quality measures, including an overall rating. We performed a correlational analysis to see if facility scores for this measure were related to the facility's overall five-star rating. PC-02 rates would be hypothesized to correlate negatively with the five-star rating.

Since the measure has been in national use, continued face validity of the measure has been determined through analysis of feedback from measure users. The Joint Commission provides a web-based application with which measure users can provide feedback regarding appropriateness of measure specifications, request clarification of specifications, and/or provide other comments pertinent to the measure. This feedback is systematically continually reviewed in order to identify trends and to identify areas of the measure specifications that require clarification or revision. Additionally, Joint Commission staff continually monitors the national literature and environment in order to assess continued validity of this measure.
As noted previously, The Joint Commission is currently performing reliability site visits this year. A component of these visits will include focus group interviews with hospital staff working with the PC measures to obtain feedback regarding the validity of the measures and suggestions for further refinement of the specifications.

ICD-9 to ICD-10 Conversion Process:
The goal was to convert ICD-9 to ICD-10 equivalent codes, consistent with the clinical intent of the original measure specifications. The Joint Commission worked with a certified coding expert throughout the conversion process. The 3M Coding Conversion Tool was utilized, including forward mapping of ICD-9 codes to ICD-10 codes as well as reverse mapping from ICD-10 to ICD-9 to ensure appropriateness. MSDRGs and instructions in the tabular index were also examined to ensure appropriate code mapping. Crosswalks comprising ICD-9 codes mapped to ICD-10 codes were created and reviewed by members of the Technical

Advisory Panel, CMS subcontractors, and performance measurement system vendors prior to being posted for a 12 month public comment period. Feedback from the field indicated that the crosswalks generally were mapped correctly. Minor modifications to the code tables were made as needed. Final code tables were published in early 2015, well in advance of the mandated date of October 1, 2015.
Perinatal Care (PC) Initial Patient Population
The PC measure set is unique in that there are two distinct Initial Patient Populations within the measure set, mothers (PC-01, PC-02, PC-03) and newborns. (PC-04, PC-05).

Subpopulation Mothers
Patients admitted to the hospital for inpatient acute care are included in the PC Mother Initial sampling group if they have: ICD-9-CM Principal or Other Diagnosis Code as defined in Appendix A, Tables 11.01, 11.02, 11.03, or 11.04, a Patient Age (Admission Date Birthdate) >= 8 years and < 65 and a Length of Stay (Discharge Date - Admission Date) ≤ 120 days
PC-02 - Cesarean Section belongs to the above population.

The data used to measure the validity of the PC measure are comprised of data from the third and fourth quarters of 2014, and the first and second quarters of 2015. 1,345 hospitals submitted 2,695,467 inpatient records for all the elected PC measures. The hospitals included in the analysis reported one year of data and had 30 or more denominator cases in the analysis period.

Measure convergent validity for PC-02 was assessed using hospitals patient level data from The Joint commission warehouse. Measure specifications, including population identification, numerator and denominator statements, exclusions, and data elements and their definitions were found to be understandable, retrievable, and relevant in previous validity testing.

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

2020 Submission

Correlations of the measure with other measures of perinatal care quality.

Table of Correlations

Measure	PC-01	PC-02	PC-05	ePC-01	ePC-05
PC-01-Elective Delivery	1				
PC-02-Cesarean Birth	0.133191771	1			
PC-05-Exclusive Breast Milk Feeding	-0.02552769	-0.28102539	1		
ePC-01-Elective Delivery	0.008935902	0.108321673	0.022811563		
ePC-05-Exclusive Breast Milk Feeding	0.040364562	-0.175224917	0.748033011	-0.457372009	1

Correlation with other measures of hospital facility quality.
A weak negative correlation was found between the facility-level PC-02 measure score and the overall quality rating ($\rho = -0.133$, $p < .0001$). This result suggests that facilities having a higher overall quality rating tend to have a lower PC-02 score, indicating higher quality in the mother's delivery.

Analysis of feedback obtained via our automated feedback system reveals slightly more than 80 submissions regarding specifications for this measure since its implementation in 2010. Predominant themes of these submissions involved questions regarding clarification of the data elements Parity and Gestational Age with respect to both definitions and the calculation of gestational age, the order of priority sources to retrieve the data and incorporation of GTPAL terminology for Parity. Additional notes for abstractors were added to the

data elements for clarification. An additional ICD-9-CM diagnosis code identifying footling breech was also added to Table 11.09 to update exclusions based on consultation with the original measure developer.

Overall descriptive statistics for sub population MOTHER:
N=1,345 hospitals
n = 2,695,467 records submitted

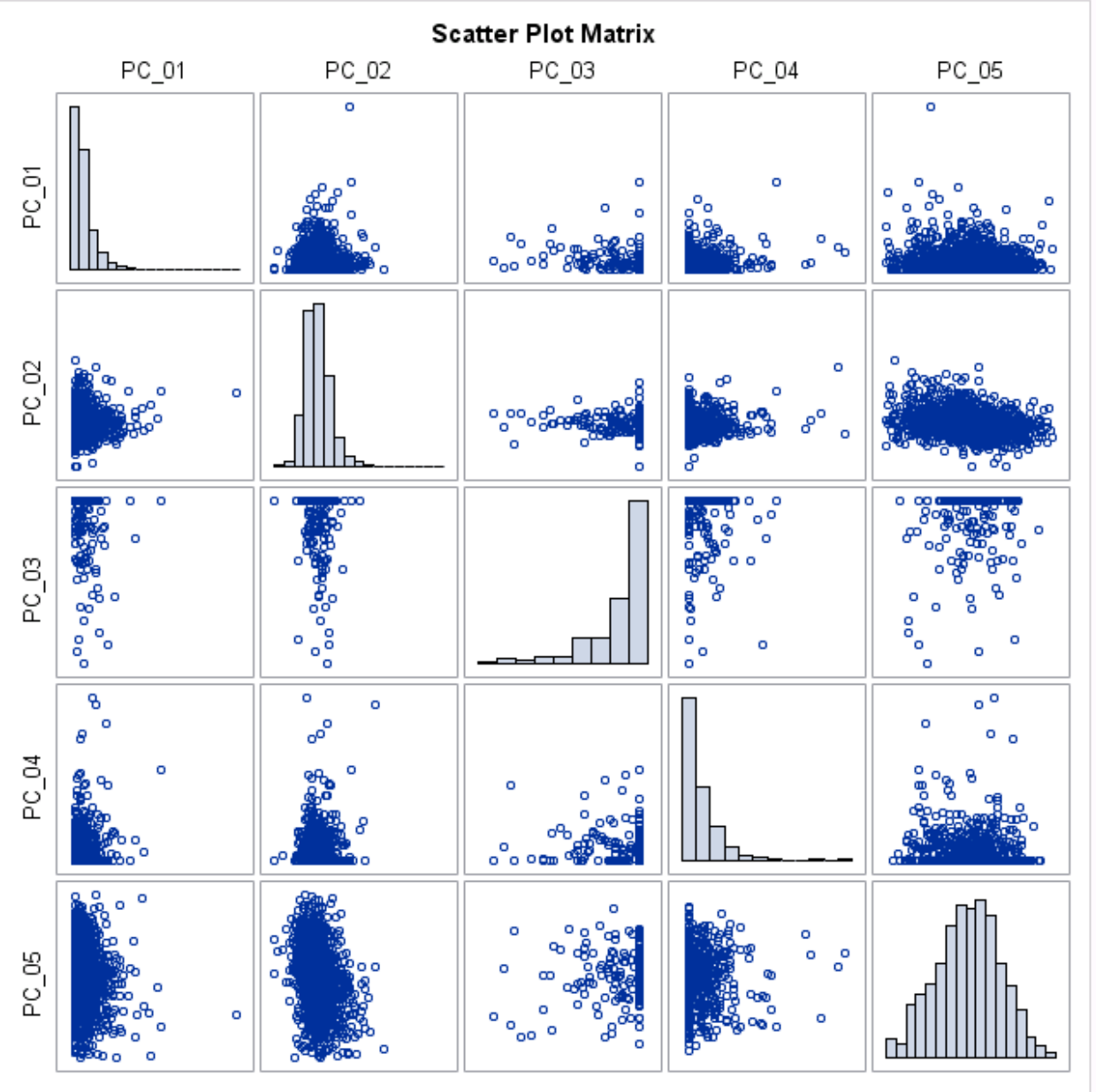
Descriptive statistics for PC-02 measure:
N=1,345 hospitals
n = 1,169,924

Min = 0%
Mean: 26.2%
Percentile 10%: 18%
Percentile 25%: 21%
Median: 25.4%
Percentile 75%: 30.4%
Percentile 90%: 36%
Max = 100%

Simple Statistics

Variable	N	Mean	Std Dev	Median	Minimum	Maximum
PC_01	1237	0.02753	0.03803	0.01734	0	0.51240
PC_02	1345	0.26287	0.07974	0.25410	0	1.00000
PC_03	162	0.97762	0.03311	0.99425	0.84615	1.00000
PC_04	523	0.05267	0.08432	0.02203	0	0.66129
PC_05	1352	0.49198	0.19284	0.50190	0.00317	1.00000

Spearman Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations					
	PC_01	PC_02	PC_03	PC_04	PC_05
PC_01	1.00000 1237	0.06843 0.0163 1231	-0.26960 0.0006 159	0.10724 0.0169 496	-0.03538 0.2137 1237
PC_02	0.06843 0.0163 1231	1.00000 1345	-0.18318 0.0196 162	0.02807 0.5218 523	-0.32009 <.0001 1343
PC_03	-0.26960 0.0006 159	-0.18318 0.0196 162	1.00000 162	-0.03117 0.7030 152	0.07729 0.3283 162
PC_04	0.10724 0.0169 496	0.02807 0.5218 523	-0.03117 0.7030 152	1.00000 523	-0.03560 0.4165 523
PC_05	-0.03538 0.2137 1237	-0.32009 <.0001 1343	0.07729 0.3283 162	-0.03560 0.4165 523	1.00000 1352



The Spearman rank-order correlation is a nonparametric measure of association based on the ranks of the data values by measure PC-02 and hospitals. We used this methodology because of the skewness of the distribution of the measure rates.

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

Correlations were found between PC-02 and the other perinatal care measures (PC-01, ePC-01, PC-05 and ePC-05) in the expected directions, as well as being correlated in the expected direction to the overall hospital five-star rating. The perinatal care measures used in this analysis are measuring different components of perinatal care and would not be expected to be more than weakly correlated since perinatal care quality is a multidimensional quantity.

The correlation of PC-02 with the other PC measures in the PC measure set indicates that the correlations

(with the exception of PC-04), although in the expected direction and statistically significant, are relatively weak. Although 90% of the hospital measure rates fall between 18 and 36%, there are still a number of hospitals with measure rates significantly greater than 36% and less than 18%, indicating that the performance of hospitals on this measure are not uniformly acceptable.

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 below.

Measure Exclusions

Exclusion	Rationale	Measure Denominator lost due to exclusion
Multiple gestations or other presentations	Table 11.09 contains diagnosis codes for a fetus in any position other than a vertex position and any gestations of two more in order to exclude these cases from the denominator, which are at higher risk with vaginal birth.	19.9%
Not a term live birth	To identify cases where the outcome of delivery is not a live birth and exclude from the denominator.	3.1%
Gestational age < 37 or UTD	The denominator population is limited to patients > 37 or more weeks of completed gestation, who are low risk with vaginal birth. Patients with UTD for gestational age typically have had no prenatal care.	9.9%
Previous live birth	To identify patients who have had a previous live birth and exclude them from the denominator, nulliparous patients are low risk with vaginal birth than compared to multiparous patients.	67.0%

Note: The exclusions presented in this table are not mutually exclusive. For example, a discharge that falls under exclusions 1 and 3 would appear in both places in this table.

We tested whether the exclusions impacted the performance score denominator.

Measure exclusions that were not derived directly from the evidence are presented below. Please note that these are population exclusions that are necessary to ensure consistency in all measures in this 4-measure set. These exclusions were analyzed for frequency of occurrence. An issue that is of great concern to users of this measure is that due to the presence of exceptions to the measure, attainment of a 0% measure rate is not possible. Because of the role of this measure in the current Joint Commission accreditation process, this is especially troubling to measure users. This concern is the basis for the non-evidence-based exclusions to these measures. Additional reasons for these population exclusions are enumerated in our response to section 2b1.1 above. The measure exclusions that were not derived directly from the evidence are as follows:

1. Patients with LOS <120 days
2. Patients less than 8 years of age or greater than or equal to 65 years of age
3. Patients enrolled in clinical trials

There were 1,169,924 admissions selected from the initial cohort. From among the 1,169,924 admissions in 1,345 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 multiple gestations and other presentations as defined in Appendix A, Table 11.09
- 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 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

Exclusion	Rationale	Measure Denominator lost due to exclusion
Multiple gestations or other presentations	Table 11.09 contains diagnosis codes for a fetus in any position other than a vertex position and any gestations of two more in order to exclude these cases from the denominator, which are at higher risk with vaginal birth.	19.9%
Not a term live birth	To identify cases where the outcome of delivery is not a live birth and exclude from the denominator.	3.1%
Gestational age < 37 or UTD	The denominator population is limited to patients > 37 or more weeks of completed gestation, who are low risk with vaginal birth. Patients with UTD for gestational age typically have had no prenatal care.	9.9%
Previous live birth	To identify patients who have had a previous live birth and exclude them from the denominator, nulliparous patients are low risk with vaginal birth than compared to multiparous patients.	67.0%

Number and percent of denominator remaining after exclusions

PC-02 Denominator		
PC-02 denominator before exclusions	PC-02 denominator after exclusions	Percent after exclusions
1,552,605	497,903	32.1%

The percentiles for the hospital percent after exclusions had the following values for the 10th, 25th, 50th, 75th and 90th percentiles respectively: 25.7%, 28.5%, 31.3%, 34.1%, and 37.9%.

N=353,671

1. Patients who have a length of stay (LOS) greater than 120 days =0%
2. Patients less than 8 years of age or greater than or equal to 65 years of age=0%
3. Patients enrolled in clinical trials =0.04%

Exclusion Subpopulation 1 – PC-02:
ICD-9-CM Principal Diagnosis Code or ICD-9-CM Other Diagnosis Codes for multiple gestations and other presentations as defined in Appendix A, Table 11.09:
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: Patients enrolled in clinical trials
Overall Occurrence n =729
Overall Occurrence Percentage: 0.06%
Minimum: 0%
10th Percentile: 0%
Median: 0%
90th Percentile: 0.03%
Maximum: 7.97%

Exclusion: Gestational Age < 37 weeks or UTD
Overall Number of Occurrences n = 113,520
Overall Occurrence Percentage: 9.7%
Minimum: 0.29%
10th Percentile: 5%
Median: 8.7%
90th Percentile: 14.8%
Maximum: 34%

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. All exclusions are necessary to ensure the construct validity of the measure and all have a clinical rationale. In the specifications, these exclusions have been incorporated into the measure definition. It should be noted that this high number is expected since it is not the population of interest as defined by this measure. Although the initial patient population is used as the basis for sampling for the measure, it is not the population of interest. Therefore, in the specifications these exclusions from the initial patient population have been incorporated into the denominator definition.

The difference between the 10th and 90th percentiles of the distribution of exclusion rates is narrow indicating that the occurrence is random and likely would not bias performance results.

It is believed that all of the exclusions should be retained for the following reasons:
Exclusion: ICD-9-CM Principal Diagnosis Code or ICD-9-CM Other Diagnosis Codes for multiple gestations and other presentations as defined in Appendix A, Table 11.09:
Rationale: *Table 11.09 contains diagnosis codes for a fetus in any position other than a vertex position and any gestations of two more in order to exclude these cases from the denominator.*

Exclusion: Patients who have a Length of stay greater than 120 days
Rationale: Included for this measure in order to harmonize with other CMS/Joint Commission aligned measures.

Exclusion: Patients with Gestational Age < 37 weeks or UTD
Rationale: The denominator population is limited to patients > 37 or more weeks of completed gestation. Patients with UTD for gestational age typically have had no prenatal care.

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.
Not applicable

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.

2020 Submission

RATIONALE

This measure is not risk-adjusted. When constructing the measure, the exclusion criteria were chosen to ensure that the target population would be healthy, term pregnancies with no pre-existing complications for the mothers, thus reducing bias due to case mix complications. Mothers more at risk for experiencing adverse outcomes were excluded from the target population. The rationale for each of the exclusions is outlined in the Exclusions section. Evidence continues to support no further risk adjustment is indicated as described below by Dr. Elliott Main, an article that is in the review process for publication.

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

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

In July 2017, the Society for Maternal-Fetal Medicine Health Policy and Advocacy Committee and Society for Maternal-Fetal Medicine Coding Committee published an expert opinion[1] suggesting additional ICD-10-CM codes that to add to the definition of low-risk birth for the purpose of cesarean birth calculation. They specifically sought diagnosis codes that represented "clinically relevant risk factors that are absolute or relative contraindications to vaginal birth." The choice of codes was not based on actual data but solely on expert opinion.

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

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

The following tables are taken from a manuscript in preparation.

Analysis of SMFM Proposed Additions to NTSV Exclusion Code Set

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

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

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

	Hospital Type					
Diagnosis Groups Proposed For Exclusion (based on ICD-10 codes)	University (main campus) (Hosp N=9) (Pt N=15,071)	AAP Level 3/4 (Hosp N=108) (Pt N=211,903)	AAP Level 2 (N=57) (Pt N=65,686)	AAP Level 1 (not Critical Access) (Hosp N=61) (Pt N=29,046)	Critical Access (Hosp N=12) (Pt N=1,684)	All Hospitals (Hosp N=238) (Pt N=308,319)
Care of Fetal anomalies	46 (3.1)	224 (1.1)	37 (0.6)	2 (0.1)	1 (0.6)	264 (0.9)
HIV	23 (1.5)	49 (0.2)	10 (0.2)	1 (0.0)	0 (0.0)	60 (0.2)
Severe	46 (3.1)	552 (2.6)	171 (2.6)	79 (2.7)	8 (4.8)	810 (2.6)

Preeclampsia						
Cardiovascular	332 (22.0)	1584 (7.5)	357 (5.4)	120 (4.1)	15 (8.9)	2076 (6.7)
Kidney HTN	5 (0.3)	35 (0.2)	4 (0.1)	1 (0.0)	0 (0.0)	40 (0.1)
Cerebral Thrombosis	0 (0.0)	1 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	2 (0.0)
Previa expanded	2 (0.1)	19 (0.1)	5 (0.1)	1 (0.0)	0 (0.0)	25 (0.1)
Low-lying placenta	31 (2.1)	307 (1.4)	91 (1.4)	45 (1.5)	2 (1.2)	445 (1.4)
Accreta	6 (0.4)	75 (0.4)	22 (0.3)	9 (0.3)	0 (0.0)	106 (0.3)
Abruption	3 (0.2)	21 (0.1)	4 (0.1)	2 (0.1)	0 (0.0)	27 (0.1)
Cord Prolapse	11 (0.7)	210 (1.0)	49 (0.7)	48 (1.7)	3 (1.8)	310 (1.0)
Vasa Previa	3 (0.2)	35 (0.2)	1 (0.0)	3 (0.1)	0 (0.0)	39 (0.1)
Any of the above	505 (33.5)	3078 (14.5)	745 (11.3)	305 (10.5)	29 (17.2)	4157 (13.5)

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

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

	Hospital Type					
Diagnosis Groups Proposed For Exclusion (based on ICD-10 codes)	University (main campus) (Hosp N=9) (Pt N=15,071)	AAP Level 3/4 (Hosp N=108) (Pt N=211,903)	AAP Level 2 (N=57) (Pt N=65,686)	AAP Level 1 (not Critical Access) (Hosp N=61) (Pt N=29,046)	Critical Access (Hosp N=12) (Pt N=1,684)	All Hospitals (Hosp N=238) (Pt N=308,319)
Care of Fetal anomalies	54.3	40.2	32.4	0.0	0.0	38.6
HIV	39.1	44.9	20.0	0.0	*No cases*	40.0
Severe Preeclampsia	34.8	47.8	55.6	50.6	37.5	49.6
Cardiovascular	28.9	32.1	35.6	33.3	26.7	32.7
Kidney HTN	20.0	25.7	75.0	0.0	*No cases*	30.0
Cerebral Thrombosis	*No cases*	100.0	0.0	*No cases*	*No cases*	50.0
Previa expanded	50.0	42.1	20.0	0.0	*No cases*	36.0
Low-lying placenta	61.3	54.4	47.3	57.8	100.0	53.5
Accreta	16.7	44.0	40.9	22.2	*No cases*	41.5
Abruption	33.3	52.4	75.0	100.0	*No cases*	59.3
Cord Prolapse	81.8	84.3	83.7	68.8	66.7	81.6
Vasa Previa	66.7	71.4	100.0	66.7	*No cases*	71.8
Any of the above	35.2	41.9	44.4	46.2	37.9	42.6

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

for these complications was relatively low.

Table 3. Revised NTSV Cesarean Delivery Rate (%) with selected major obstetric complications excluded (among NTSV PC-02 population)

	Hospital Type					
Diagnosis Groups Proposed For Exclusion (based on ICD-10 codes)	University (main campus) (Hosp N=9) (Pt N=15,071)	AAP Level 3/4 (Hosp N=108) (Pt N=211,903)	AAP Level 2 (N=57) (Pt N=65,686)	AAP Level 1 (not Critical Access) (Hosp N=61) (Pt N=29,046)	Critical Access (Hosp N=12) (Pt N=1,684)	All Hospitals (Hosp N=238) (Pt N=308,319)
Baseline NTSV (PC-02)	22.7	24.6	24.0	25.9	21.7	24.6
Care of Fetal anomalies	22.6	24.6	23.9	25.9	21.7	24.5
HIV	22.7	24.6	24.0	25.9	21.7	24.6
Severe Preeclampsia	22.7	24.5	23.9	25.8	21.6	24.5
Cardiovascular	22.6	24.5	23.9	25.9	21.6	24.5
Kidney HTN	22.7	24.6	23.9	25.9	21.7	24.6
Cerebral Thrombosis	22.7	24.6	24.0	25.9	21.7	24.6
Previa expanded	22.7	24.6	24.0	25.9	21.7	24.6
Low-lying placenta	22.6	24.5	23.9	25.8	21.6	24.5
Accreta	22.7	24.6	23.9	25.9	21.7	24.6
Abruption	22.7	24.6	23.9	25.9	21.7	24.6
Cord Prolapse	22.7	24.5	23.9	25.8	21.6	24.5
Vasa Previa	22.7	24.6	24.0	25.9	21.7	24.6
Any of the above	22.3	24.3	23.7	25.7	21.4	24.3

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

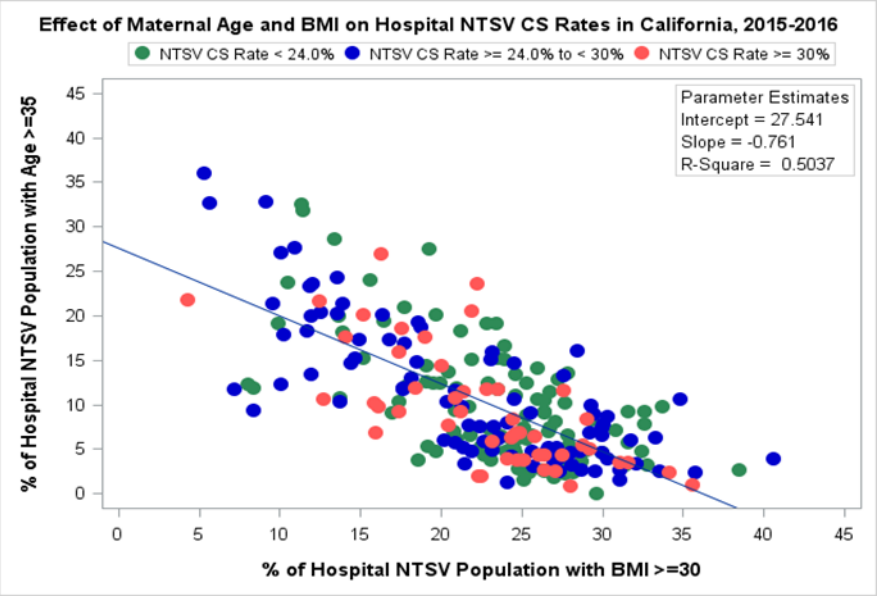
2. CMQCC Analysis of Effects of Maternal Age and BMI on NTSV Cesarean rate

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

Descriptive Approach: In Figure 1 we have graphed the proportion of the hospital's birthing population that has advanced maternal age (≥35 years) versus the proportion of the hospital's population that has a pre-pregnancy BMI >30 for 242 California hospitals with an average of ≥100 annual births continually open from 2015-2016. A moderate correlation between age and BMI is noted. The hospital dots are color coded

by their NTSV rate: green for <24%, blue for 24-30% and red for >30%. There are two notable observations: (1) green and red dots are widely distributed through the graph; and (2) for every Age/BMI intercept with a red dot there are multiple green dots nearby with similar Age/BMI populations. This would support the conclusion that provider/nursing practice(s) is the main driver for the variation in care noted for age/BMI and lack of need for adjustment.

Figure 1. Overlap of Age and BMI populations for high and low NTSV Cesarean rate hospitals



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

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

After the submission of this form and before the standing committee meeting at which this measure was discussed the Joint Commission's Perinatal Care Technical Advisory Panel recommended using the simple cesarean birth rates without further risk adjustment. The decision to remove all risk-adjustment from this measure was made based on analysis of data on this measure received by The Joint Commission which indicates that age is only a weak predictor of outcome and that age standardization could potentially distort

the age-standardized measure rates for hospitals with small sample sizes. Additionally, the Technical Advisory Panel considered evidence from two recent studies [1], [2] when making the recommendation to remove age standardization from the measure. Therefore, effective with discharges beginning July 1, 2016, The Joint Commission has removed all risk adjustments.

[1] Caceres IA, Arcaya M, Declercq E, Belanoff CM, Janakiraman V, et al. (2013) Hospital Differences in Cesarean Deliveries in Massachusetts (US) 2004–2006:The Case against Case-Mix Artifact. PLoS ONE 8(3): e57817. doi:10.1371/journal.pone.0057817

[2] Main E. (2014) Nuliparous, Term, Singleton, Vertex (NTSV) Cesarean Birth Rates: extreme hospital variation is not changed by adjustment for case-mix. Oral Presentation: Pacific Coast Obstetrics and Gynecology Society

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

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

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*)

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 calculates 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.

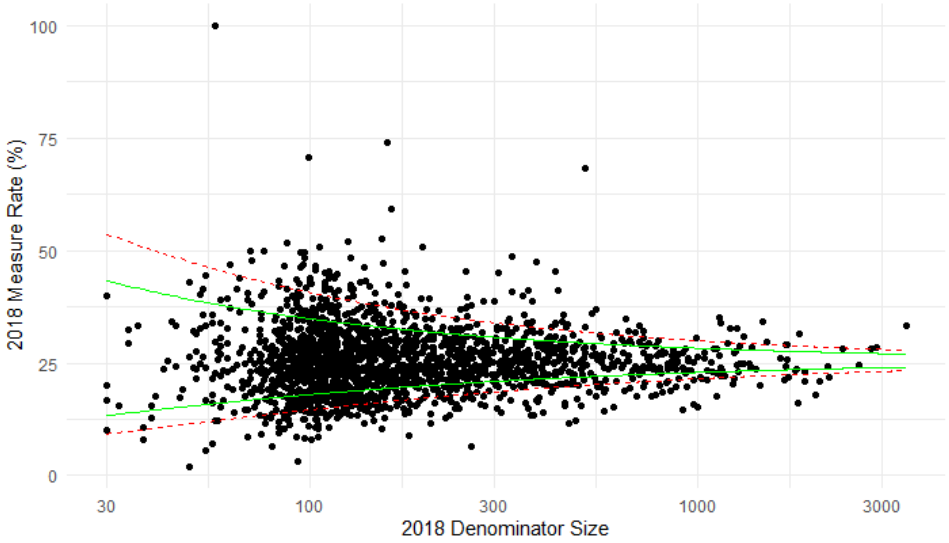
There are two components to the target analysis methodology used at The Joint Commission. Given the national average for a performance measure, a target range is constructed. Using generalized linear mixed models' methodology (also known as hierarchical models), a predicted estimate of an HCO's performance, 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.

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, 289 hospitals were identified as outliers with rates beyond the 2 standard deviation upper limit and 122 hospitals were identified as outliers with rates beyond the 3 standard deviation limits. For example, the upper limit of a 95% confidence interval for a hospital with the median denominator size of 142 is 33.3%, and 38.1% for a 99.9% confidence interval.

Funnel Plot for PC-02:



PC-02 Distribution of Rates
2018 Data:
Scores on this measure: N=1936, Mean 25.7%, SD 7.6%
10th Percentile= 16.9%
25th Percentile= 20.8%
50th Percentile= 25.0%
75th Percentile= 29.6%
90th Percentile= 34.8%

PC-02 Distribution of Outliers
2011 1st Quarter Data:
Scores on this measure: N=160, Mean 26.7%, SD 0.12953
10th Percentile= 14%
25th Percentile= 19.4%
50th Percentile=26%
75th Percentile= 32.5%
90th Percentile= 40
159 (100%) Neutral – results not significantly different from target range

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 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.

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

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

2b5.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)
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
2020 Submission
The measure has been collected since 2011 and hospitals transmitting data with missing data on any of the critical data elements are not accepted.

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 non-responders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)
Not applicable

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*)
Not applicable

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 non-responders) 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*)
Not applicable

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 by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition, Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

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**.

Some data elements are in defined fields in electronic sources

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).

The Joint Commission recognizes that not all hospitals currently have the capacity to abstract the electronic version of this measure, so continues to offer this chart abstracted version which allows for data capture from unstructured data fields. All data elements needed to compute the PC-02 performance measure score were retooled for capture from electronic sources in 2016. Specifications were updated in 2019 based on testing and current eCQM standards. The measure will be available to hospitals in 2020 for data collection to meet Joint Commission accreditation requirements for eCQM submission.

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:

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.

At the present time, hospitals using this performance measure generally collect measure data via manual review of the EMR, data derived from vital records reports received from state or local departments of public health, delivery logs or clinical information systems or a combination. Collected data are submitted to The Joint Commission on a quarterly basis, as described previously. Specifications for this measure are freely available to anyone who wishes to use the measure. Feedback from hospitals using this measure indicates that required data elements are generally available in the medical record, and measure specifications are robust and easy to understand. As feedback from measure users has indicated the need for clarification or revision of measure specifications, this has taken place.

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

There are no fees 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	Regulatory and Accreditation Programs Hospital Accreditation Program http://jointcommission.org Hospital Accreditation Program http://jointcommission.org Quality Improvement (Internal to the specific organization) Perinatal Care Certification http://www.jointcommission.org/certification/perinatal_care_certification.aspx

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

- 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; 2005 Joint Commission-accredited hospitals reported PC-02 in 2018 (67% of Joint Commission-accredited hospitals), 1986 reported one or more denominator cases (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.

Name of program and sponsor: Quality Check®- The Joint Commission

Public Reporting of PC-02 will begin July 2020

Quality Check®

http://www.qualitycheck.org/consumer/searchQCR.aspx

- Purpose: A public website that allows consumers to: search for accredited and certified organizations by city and state, by name or by zip code (up to 250 miles); find organizations by type of service provided within a geographic area; download free hospital performance measure results; and, print a list of Joint Commission certified disease-specific care programs and health care staffing firms.
- Geographic area and number and percentage of accountable entities and patients included: Nationwide; 3895 Joint Commission-accredited hospitals (2019)

Name of program and sponsor: Perinatal Care Certification- The Joint Commission

- Purpose: A certification program that recognizes hospitals that have achieved integrated, coordinated, patient-centered care for clinically uncomplicated pregnancies and births.
- Geographic area and number and percentage of accountable entities and patients included: Nationwide; 65 Joint Commission-accredited hospitals (2018)

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?*)

The Joint Commission will begin publicly reporting hospitals with consistently high cesarean birth rates on Quality Check® July 1, 2020, using data reported by hospitals.

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.*)

The Cesarean Birth measure PC-02 measures the rates of cesarean births amongst a subset of the general obstetric population of low-risk women having their first birth with a term, singleton baby in a vertex position (NTSV).

The Joint Commission will use data reported by hospitals during the calendar years 2018 and 2019, along with the following three criteria to determine a hospital's PC-02 rating:

1. =30 cases reported in both years

2. PC-02 rate >30% for the current year

3. Overall twenty-four-month aggregate PC-02 rate >30% (see note below)

Note: 2018 and 2019 data will be used for the initial release. Moving forward the overall twenty-four-month aggregate rate will be calculated from a rolling eight calendar quarters and refreshed on Quality Check biannually in July and January.

Hospitals will be identified on Quality Check with either a plus (+) or minus (-) symbol for the PC-02 measure.

- The plus (+) symbol will signify the hospital has an acceptable rate.

• A minus (-) symbol will signify the hospital's rate is consistently high and has a large enough sample size to make this determination.

Avoiding Unhealthy Consequences

For those hospitals identified as having high rates (-), The Joint Commission will also show those hospitals' actual 2019 PC-02 rates. Hospitals with acceptable rates (+) will not have the actual PC-02 rates reported. The Joint Commission believes hospitals should work to reduce unnecessary cesarean births; however, it does not want to differentiate between groups of hospitals whose rates are in the acceptable range. Lower is not always better in these cases, and The Joint Commission does not want to encourage inappropriately low rates that may be unsafe to patients.

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.

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 is aggregated 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.

Most statistical questions on this measure were regarding how this measure was to be publicly reported in 2020. There was strong support for the public reporting of this measure from multiple stakeholders.

Queries submitted via the automated feedback system have decreased significantly for the early elective delivery measure in the past three years.

Correction

The statement "Queries submitted via the automated feedback system have decreased significantly for the measure in the past three years." does apply for PC-02.

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 workgroup to confirm the need for revision. Additionally, The Joint Commission engages a Technical Advisory Panel (TAP) for review and/or approval of updates which may require their 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.

Unintended Consequence:

Patients who did not receive prenatal care were inappropriately included in the measure denominator, as the gestational age data element was abstracted as unable to be determined (UTD).

Mitigating Action: In order to avoid penalizing hospitals, cases with UTD were removed from the measure population.

Unintended Consequence:

Some hospitals have reported higher rates due to small denominator populations as a result of sampling.

Mitigating Action:

Vital Records reports, delivery logs and clinical information systems were added as acceptable data sources to help hospitals identify all cases with =>37 weeks gestation, so that 100% of these cases could be reviewed to increase the denominator population size.

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

Not applicable

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

No

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

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

5a. Harmonization of Related Measures

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

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

No

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

Not Applicable

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 timeframe. After implementation, minor revisions, acknowledged by TAP representatives, were made to improve clarity. Hospital feedback will be reviewed during the reliability testing phase of the project to assist the TAP in making the final measure recommendations.

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2010

Ad.3 Month and Year of most recent revision: 10, 2015

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

Ad.5 When is the next scheduled review/update for this measure? 01, 2020

Ad.6 Copyright statement: No royalty or use fee is required for copying or reprinting this manual, but the following are required as a condition of usage: 1) disclosure that the Specifications Manual is periodically updated, and that the version being copied or reprinted may not be up-to-date when used unless the copier or printer has verified the version to be up-to-date and affirms that, and 2) users participating in Joint

Commission accreditation, including vendors, are required to update their software and associated documentation based on the published manual production timelines.

Ad.7 Disclaimers:

Ad.8 Additional Information/Comments: