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 #: 0716

Corresponding Measures:

De.2. Measure Title: Unexpected Newborn Complications in Term Infants

Co.1.1. Measure Steward: California Maternal Quality Care Collaborative

De.3. Brief Description of Measure: This is a hospital level performance score reported as the percent of infants with Unexpected Newborn Complications among full term newborns with no preexisting conditions, typically calculated per year.

1b.1. Developer Rationale: The most important childbirth outcome for families is bringing home a healthy baby. While there have been measures developed to assess clinical practices and outcomes in preterm infants, there are a lack of metrics that assess the health outcomes of term infants who represent over 90% of all births.

The Unexpected Complications in Term Newborns metric addresses this gap and measures adverse outcomes resulting in severe or moderate morbidity in otherwise healthy term infants without preexisting conditions. Importantly, this metric also serves as a balancing measure for other NQF endorsed maternal measures such as NTSV Cesarean rates, third and fourth degree lacerations, episiotomies and early elective delivery rates. The purpose of a balancing measure to is guard against any unanticipated or unintended consequences of quality improvement activities for these measures.

S.4. Numerator Statement: Numerator: The numerator is divided into two categories: Severe complications and moderate complications.

Severe complications include neonatal death, transfer to another hospital for higher level of care, , severe birth injuries such as intracranial hemorrhage or nerve injury, neurologic damage, severe respiratory and infectious complications such as sepsis. Parents of such babies may often worry about short or long term infant outcomes.

Moderate complications include diagnoses or procedures that raise concern but at a lower level than the list for severe (e.g. use of CPAP or bone fracture). For inclusion in the numerator, most require an infant length of stay that exceeds that of the mother, validating that these are indeed significant complications. Examples include less severe respiratory complications (e.g. Transient Tachypnea of the Newborn), or infections with a longer length of stay not including sepsis. As a "safety net" to capture cases who were under-coded, the numerator also includes infants who have a prolonged length of stay of over 5 days to capture the "seemingly normal" infants with neither any form of jaundice nor a social reason for staying in the hospital (e.g. family disruption or adoption).

S.6. Denominator Statement: The denominator is comprised of singleton, live born babies who are at least 37.0 weeks of gestation, and over 2500g in birth weight. The denominator excludes most serious fetal conditions that are "preexisting" (present before labor), including prematurity, multiple gestations, poor fetal growth, congenital malformations, genetic disorders, other specified fetal and maternal conditions and infants exposed to maternal drug use in-utero. The final denominator population consists of babies who are expected to do well following labor and delivery and go home routinely with their mothers.

S.8. Denominator Exclusions: a) Babies not born in hospitals are excluded as this is a hospital quality performance measure

b) Babies who are part of multiple gestation pregnancies are excluded.

c) Premature infants (babies born before 37 weeks gestational age) are excluded

d) Low birth weight babies (<=2500g) are excluded

e) Babies with congenital malformations and genetic diseases are excluded

f) Babies with pre-existing fetal conditions such as IUGR are excluded

g) Babies who were exposed to maternal drug use in-utero are excluded

De.1. Measure Type: Outcome

S.17. Data Source: Claims

S.20. Level of Analysis: Facility, Integrated Delivery System, Population : Regional and State

IF Endorsement Maintenance – Original Endorsement Date: Jan 17, 2011 **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? This is not a paired or grouped measure but we anticipate that this measure will serve as a balancing measure for other NQF endorsed maternal measures such as NTSV Cesarean rates, third and fourth-degree lacerations, episiotomies and early elective delivery rates. The purpose of a balancing measure is to guard against any unanticipated or unintended consequences of quality improvement activities for these measures. Most families would value having a low chance of unexpected newborn complications and low-medium rates of obstetric procedures (such as Cesarean births and operative vaginal deliveries).

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: HEALTHY TERM FETUS: no preexisting conditions and expected to go home with no complications >>exposed to>> QUALITY OF CARE PROCESSES during labor management, delivery or neonatal care such as (but not limited to): use of forceps/vacuums, poor management of intrapartum fetal heart tracings, poor management of antibiotic prophylaxis for prevention of neonatal infection, etc. >>resulting in>> ADVERSE OUTCOME e.g., severe or moderate morbidities in the newborn including death, transfer to a higher level of care, birth injuries, infections, respiratory complications, neurological complications, shock/resuscitation or prolonged neonatal length of stay.
- Evidence from the previous submission was not guideline-based.

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 noted five studies in the last 18 months have used Unexpected Newborn Complications as either a key outcome or important balancing measure during studies focused on improving obstetric practice and offer comparisons to other simultaneously collected neonatal outcome measures.
  - The studies support the correlation between total Unexpected Newborn Complications and Severe Unexpected Newborn Complications with other commonly used neonatal outcome metrics (e.g., Apgar scores, birth injuries, and umbilical artery base deficit).
  - Several of the studies showed improvements in Unexpected Newborn Complications following large-scale quality improvement/safety initiatives, thus illustrating the ability to influence performance on the measure.

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.

The developer provided 2017 performance data from 225 California hospitals (>200 cases; 372,139 eligible births), as follows:

- Distribution of the rate of unexpected newborn complications

Mean	Std Deviation	Median	Q1	Q3	Min	Max
3.19	1.88	2.74	1.84	4.07	0.21	11.21

- The developer also provided the distribution of these data by deciles.
- Statewide performance (all hospitals), rate per 1,000 births over time:

2013	32.3
2014	32.3
2015	30.4
2016	29.7
2017	29.3
2018	30.0
2019	29.1

Disparities

- The developer provided performance data from 225 California hospitals (>200 cases; 372,139 eligible births) stratified by maternal race/ethnicity, as follows:

Characteristic	Numerator N=11,547	Denominator N=372,139	Unexpected newborn complication rate (%)
Maternal race/ethnicity			
White	3,283	102,851	3.19
Black	680	18,635	3.65
Asian	1,541	59,267	2.60
Hispanic	5,556	177,520	3.13
Other	108	3,271	3.30
Missing	379	10,595	3.58

- Other stratification variables provided included Neonatal Gender; Insurance Status (Medi-Cal, Private, Self-pay; Geographic Location of Hospital, AAP Neonatal Level of Care; Geographic Location of Hospitals; Hospital Ownership.

Question for the Committee:

- Is there a gap in care and/or disparities 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:

** There is significant and demonstrable evidence to support the measure focus.

** Complications of term newborns is an outcome measure. It does not need to be risk adjusted as it is stratified instead.

** The developer submitted additional studies to support this measure.

** There is direct correlation. New evidence was submitted.

1b. Performance Gap

Comments:

** The performance measure was limited by the lack of adequate stratification of the hospitals and the level of maternal and neonatal care they are capable of delivering. There are studies to suggest that a minimal number of deliveries per month are required for the staff to be comfortable and maintain their skills. This was not taken into consideration in this draft. There was a suggestion that some Level I hospitals would benefit from education lleved and more training of the staff to increase their comfort level

** yes, substantial differences between hospitals, some differences associated with hospital characteristics

** The developer submitted data that show a wide variation in performance in hospitals reporting this measure. These gaps in care warrant a national performance measure.

** There is a gap in performance and warrants a continued measure.

1b. Disparities

Comments:

** There were no SES data or variables included for risk adjustment, which is a major flaw.

** There are differences by race/ethnicity. Unclear what these mean

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

** Yes, and disparities are seen with race/ethnicity

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? ☒ **Yes** ☐ **No**

Evaluators: NQF Scientific Methods Panel Subgroup

Methods Panel Review (Combined)

Methods Panel Evaluation Summary:

- This measure was reviewed by the Scientific Methods Panel. A summary of the measure and the Panel discussion is provided below.

Scientific Methods Panel Votes: Measure passes

- Reliability: H-5; M-3; L-0; I-1
- Validity: H-3; M-4; L-1; I-1

Reliability

- Score-level reliability testing (beta-binomial) was performed on a 2017 dataset (225 hospitals, which excluded 13 hospitals with fewer than 200 cases; 372,139 singleton term newborns without preexisting conditions).
- A score of 0.90 was achieved, which the developer reports as very good—i.e., differences due to performance and not measurement error (Adams, 2009). A decile analysis also was performed and yielded scores >0.7 in 9/10 deciles and >0.9 in 7/10 deciles.
- The developer concluded that reliability testing results demonstrate that variation in scores is caused by real differences in performance across the hospitals and is not due to measurement error.

Validity

- Four empirical score-level validity tests were performed, with the samples varying depending on the test.
  - Construct validity testing of this measures was performed against NICU admission; hospital cost; and LOS: a patient-level analysis; a hospital-level analysis; a comparison of ICD-9 vs. ICD-10 coding periods; and a Pearson's Correlation Coefficient between this measure and NICU admissions based on chart review data.
  - The developer reported:
    - "Exceptionally large average differences between babies with and without Unexpected Newborn Complications for both LOS and cost."
    - Hospital average newborn LOS and cost were positively associated with hospital rate of Unexpected Newborn Complications (coefficients of 0.41 and 0.37, respectively)
    - The rate was stable from the ICD-9 to ICD-10 period.
    - The correlation coefficient of Unexpected Newborn Complications and NICU admission was 0.64.

Questions for the Committee regarding reliability:

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

Questions for the Committee regarding validity:

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

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

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

2a1. Reliability – Specifications

Comments:

** I am concerned about the inadequate risk/case-mix adjustments that were done. I am also concerned about LOS and Cost data being used as a proxy outcome measure for quality and to define an unanticipated complication.

** Algorithm makes sense. Reliability scores are quite high. Research demonstrating that is performs better than alternative measures (NICU admission). There are concerns that it could be affected by coding intensity differences rather than health states. Measure is based on administrative data so should be practical to use and implement on a wide scale.

** It is concerning that there are so many exclusions and that systems may not be set up to accurately report this measure.

** High reliability in the differences in the scores. No concerns with consistent implementation.

2a2. Reliability – Testing

Comments:

** Without a deep dive into the "social determinants" of health factors - the reliability is limited.

** I would like to see what proportion of the numerator is dominated by prolonged length of stay rather than a direct health outcome

** It is concerning that the initial submission showed measure reliability scores by hospital varied from .99 to .53. Is there a plan to improve reliability?

** No concerns.

2b1. Validity –Testing

Comments:

** I have the same concerns that I have with the reliability-the lack of a deep dive into SES and Social Determinants variables and data are problematic.

** no

** No concerns.

** No concerns.

2b2-3. Exclusions/Risk Adjustment

Comments:

** Risk adjustment was not appropriately developed and tested.

** exclusions are appropriate. Case mix adjustment arguable not needed as measure is stratified by design to term babies without underlying conditions at birth.

** I agree with the developer's rationale for not including risk adjustments.

** Appropriate exclusions are identified. Measure is not risk-adjusted but is risk-stratified using exclusions. Many social risk factors are not identified as exclusions, but may help evaluate various opportunities for improvement.

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

Comments:

** Threats to validity also include the fact that only claims data is used as a data source. Other data sources including, not exclusively, Health Risk Assessment data would be very helpful and would add statistical rigor.

** no

** No concerns.

** Differences in performance correlate with real differences in performance. No concerns with missing data.

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 noted the following:

- Coding/abstraction is performed by someone other than person obtaining original information.
- All data elements are in defined fields in a combination of electronic sources.
- It is exploring with The Joint Commission the feasibility of this measure being developed as an eCQM, with the primary barrier being bandwidth to undertake and resources.

The developer identified two coding practice issues related to implementation that needed to be addressed in the specifications:

- Over and under coding: Coding practices vary for some ICD-10 codes, with some hospitals being "over exuberant" in their coding and others clearly under-coding existing complications. To address coding variance, the new specifications attempt to balance this issue by requiring that many codes for Moderate Complications additionally have an infant length of stay (LOS) that exceeds the typical maternal postpartum LOS (>2 days for a vaginal birth and >4 days for a cesarean birth). This requirement significantly reduces the number of infants identified, but validates that these babies had significant morbidity. Conversely, some babies had very long neonatal LOS without any codes to account for it, suggesting the possibility of under-coding. The developer's expert panel identified two categories of prolonged neonatal LOS that were not medically serious and could be excluded from this consideration—neonatal jaundice typically treated with Bili-Lights and social disruption for homelessness or foster care. The developer found that a number of babies with septicemia had short LOS indicating that it was not likely severe; therefore it added a requirement for a length of stay of at least five days to be included among Severe Complications.
- Coding related to billing: The developer noted that in some higher-level hospitals, coding practices varied from standard to achieve billing targets. The most common example was noted in tertiary facilities, where the code for CPAP was routinely used for bag and mask resuscitation in the delivery room even when used for less than one minute. To address this issue, the codes for CPAP were moved to the category that required a postpartum LOS that was longer than the typical LOS (as noted above).

Question for the Committee:

- Does the Committee have concerns about the feasibility of this measure?

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

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

Comments:

** The data collection strategy should be more comprehensive and include many more sources that are available.

** no concerns, good feasibilities

** It would seem that this measure could be troubled by a plethora of coding issues. If I understand the narrative correctly, hospital-level data are analyzed by a third party and reported back to each hospital. Are

third party data analyzers assessing reliability between systems?

** Variations in coding practices can result in variations in the perceived improvement opportunities.

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

Public reporting (planned):

- The Joint Commission (data collection began in January 2020, with public reporting intended for later in 2020] and regulatory/accreditation)

Accountability:

- California Maternal Quality Care Collaborative (public health/disease surveillance)
- BC BS Blue Distinction for Maternity Care (professional certification/recognition)

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 the following:

- The developer provides monthly updates to all member hospitals, including benchmarking against like size and NICU levels, county, and perinatal region.
- The developer and The Joint Commission have several education channels: webinars, user support, and written summaries.
- Feedback to the developer is obtained via written help tickets, user groups, and forums; a mechanism for feedback to the Joint Commission also is in place.
    - Feedback to address questions has been incorporated into FAQs, including addressing the following: Is there a target rate? How to use the measure? Do hospitals caring for higher risk patients have higher rates of UNC? Is a high UNC rate due to coding or care? Will a case that has been transferred be counted for the delivery or the receiving hospital? Are extramural deliveries excluded from UNC? How are transfers of newborns with known anomalies counted; what about a transfer of a newborn at the request of an insurer?

- As previously noted, feedback has been instrumental in adjusting the coding. Feedback on the categorization of "transfer to higher level of care" as an UNC criteria also has been an important issue because it is a driver of UNC for small hospitals and represents both a separation of baby from family, as well as significant complications that are typically poorly coded by the lower level facility. The measure was not modified for this feedback but further feedback and QI efforts were initiated.

Additional Feedback:

- Not reviewed by the Measure Applications Partnership

***Questions for the Committee**:*

- Can the performance results be used to further the goal of high-quality, efficient healthcare?
- Does the Committee wish to discuss with the developer the specific timing of public reporting by the Joint Commission? by CHART/Cal Hospital Compare?

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 developer provided annual performance rates for the period 2013-2019, which fluctuate, but show improvement during this timeframe.
- The developer also stated this measure was an important balancing measure for a large scale quality improvement collaborative to reduce primary cesarean birth. Unexpected Newborn Complications did not worsen when Cesarean rates were reduced by 24%, but actually improved by 8% (Main, et al, 2019). Initial data from a statewide effort similarly demonstrate this finding.

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 no unintended consequences.
- The developer noted three unexpected findings:
  - Wide variation exists among Level 1 nurseries transferring out term babies without any prior conditions—from 0.5% to 6% among California hospitals and in a similar study in Northern New England (VT, NH and ME). In both settings, follow-up interviews identified the practice was not due to patient characteristics, but rather to staff level of experience and confidence caring for babies with mild respiratory complications (primarily the unease for even a few hours of observation for transient tachypnea), and separated the baby from the family without clear need. The developer reported the quality collaboratives identified this as an education and training opportunity.
  - One additional issue has appeared largely in California hospitals that sub-contracted their NICUs to a Children´s Hospital. As the Children´s Hospital has a separate licensure, all admissions to the its NICU, even for minor reasons, are treated technically as a transfer meeting an UNC criteria. The developer noted that because it is important for the obstetric hospital to understand the

outcomes of its infants, it has worked with the joint operations to allow for the OB facility to report all of the ICD10 codes of babies it has have "transferred" to the in-facility NICU.
    - The developer indicated it did not expect that as the California NTSV cesarean rate declined, Severe UNC rate also would improve.

Potential harms

- None reported by the developer.

Additional Feedback:

- Not reviewed by the Measure Applications Partnership

***Questions for the Committee**:*

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

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

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

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

Comments:

** The measure as I know it is not being reported publicly.

** Measure is being used in three states over multiple years.

** Use of this measure has been relatively limited. Did TJC begin data collection in January 2020 as planned? Is public reporting still planned for later in 2020?

** Reported to TJC and used in payor recognition programs. Those being measured have the ability to provide feedback.

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

Comments:

** There is a danger that potential gaps in care or early interventions that could improve outcomes will be overlooked because of the lack of risk/case mix adjustment and the exclusion of factors that impact outcome - e.g. the number of deliveries per year - which impact the staffing and skill maintenance level of the staff.

** measure is best used as a balancing measure in conjunction with other measures such as those around cesarean delivery or third and fourth degree tears.

** A measure to balance and allay fears about efforts to reduce Cesarean sections and increase VBACs is warranted. Hospitals can use this measure to assess factors contributing to high scores and develop strategies to reduce UNCs.

** Can be used as a balancing measure for NTSV C-S rates. no unintended consequences are noted with this measure.

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:

** There are no obvious related or competing measures.

** no

** No concerns

** None identified.

Public and Member Comments

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

- **Of the 0 NQF members who have submitted a support/non-support choice:**
  - 0 support the measure
  - 0 do not support the measure

Combined Methods Panel Scientific Acceptability Evaluation

Measure Number: 0716

Measure Title: Unexpected Newborn Complications in Term Infants

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**
Panel Member #8: (regional, integrated deliver system, state)

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_0716" 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.**

   Panel Member #1: I found the measure specifications very clear.

   Panel Member #2: No concern

   Panel Member #3: No concerns

   Panel Member #4: Generally, specifications are well specified. Only concern is the MIF, S.5 (numerator detail) "transfer to a higher level of care" is not defined.

   Panel Member #5: The new measure has undergone detailed reconsideration of its specification, prompted by the move to ICD10 but encompassing additional changes for coherence. These changes highlight what may be considered shortcomings in the previous specification. I would suggest that this measure undergo revision in 1-2 years once additional experience with the specifications has been gained.

   Panel Member #7: None

   Panel Member #8: No

   Panel Member #9: No concerns.

   Panel Member #10: none

RELIABILITY: TESTING

Submission document: "MIF_0716" 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

   Panel Member #4: NA – score level testing conducted

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

   Submission document: Testing attachment, section 2a2.2

   Panel Member #1: The STN analyses was appropriately used based on the Adams tutorial for score level reliability.

   Panel Member #2: Adams method for reliability of provider profiling. 225 hospitals across CA..beta binomial model for between and within hospital variance

   Panel Member #3: Conducted signal-to-noise analysis, which is an appropriate method

   Panel Member #4: Testing methods are adequate regarding measure testing of. Only concern is the sample is limited to 1 state: CA. It's unclear the number of hospitals that were used as it says "in California who had over 200 cases". So, it's not clear whether: 225 hospitals had 200 or more cases, or they began with 225 hospitals and a subset were used that had at least 200 cases. Given the fact 2a2.3 refers to 225 hospitals, assume it's the former.

"Reliability is estimated using a -binomial model, which is appropriate for measuring the reliability of the UNC metric by hospital reliability score. The strategy involves fitting a beta-the performance metric results. Two parameters (alpha and beta) that define the beta-binomial distribution are generated from the model. From these parameters, the "hospital variance" was 0.9produced. Next, the was generated based on the proportion of affirmative answers. Analyzing the between hospital variance and the within hospital variance generates the reliability for each hospital site." [p7].

Panel Member #7: Adequate

Panel Member #8: A signal to noise ratio was calculated for 225 hospitals in California who had over 200 cases meeting the denominator inclusion criteria. The main hospital reliability score was 0.9

Panel Member #9: The beta-binomial methods was appropriately used, but only for hospital with more than 200 deliveries.

Panel Member #10: SNR based on the beta binomial model. Overall reliability is excellent (0.90). Looked at distribution of reliability by decile (presumably deciles here were based on actual reliability rating. This is not ideal. Should have instead grouped hospitals based on deciles of case volume, and then evaluated reliability to determine what the volume threshold is for reliability to be equal to or greater than 0.70. This approach is suggested but certainly not required in the NQF white paper.)

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

Submission document: Testing attachment, section 2a2.3

Panel Member #1: Reliability of measure score level was supported (mean=0.9) only for hospitals that have at least 200 cases for the measurement period, which is the developers' recommendation for this measure. This should be specified in the final measure endorsement reports.

Panel Member #2: Reliability = 0.90

Panel Member #3: The facility-level results were good, with a mean reliability score of 0.90. No results were provided for Integrated Delivery System or Regional/State Populations

Panel Member #4: While the is good, there is an issue with performance smaller hospitals per Table 2 (beginning p8). Of concern among the 225 hospitals: -4 (1.8%) hospitals with a reliability score of 0.59 or less

-9 (4%) hospitals with a reliability score between 0.60 - 0.69 or less

Thus, arguably nearly 6% of hospitals that qualify to be rated had poor reliability. It suggests to increase the minimum denominator threshold to qualify to be rated. Regarding these 13 hospitals, the mean denominator was 350 cases with a range from 220 – 613.

"…mean reliability score for the 225 hospitals was 0.90…." [p7

"Table 1. Distribution of Unexpected Newborn Complication Reliability Score, summarized by Hospital Deciles." [p7]

"Table 2. Unexpected Newborn Complication Reliability Score, by hospital" [p8]

Panel Member #7: Adequate however about 6% of hospitals with >200 cases fell below the recommended reliability cutoff of .70.

Panel Member #8: No Concerns

Panel Member #9: The results demonstrated a high level of reliability for hospital with more than 200 deliveries, but the measure is not restricted hospital with this number of deliveries.

Panel Member #10: SNR based on the beta binomial model. Overall reliability is excellent (0.90). Looked at distribution of reliability by decile (presumably deciles here were based on actual reliability rating. This is not ideal. Should have instead grouped hospitals based on deciles of case volume, and then evaluated

reliability to determine what the volume threshold is for reliability to be equal to or greater than 0.70. This approach is suggested but certainly not required in the NQF white paper.)

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)

   Panel Member #7: Although the magnitude of between hospital differences (vs. error) is difficult to assess at all ranges of the distribution

   Panel Member #9: (for hospitals with at least 200 deliveries)

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.**

    Panel Member #1: Methods and interpretation seem appropriate, and reliability of measure scores were high, therefor the 'high' rating. However, as noted above and acknowledged by the developers, results pertain only to hospitals with at least 200 cases.

    Panel Member #3: If measure is going to be specified for Integrated Delivery Systems or Populations: Regional or State, reliability testing should be provided for that level of analysis.

    Panel Member #4: Response to Q7: While the mean reliability score of 0.90 is good, there is an issue with performance for smaller hospitals per Table 2 (beginning p8). Of concern among the 225 hospitals:

    -4 (1.8%) hospitals with a reliability score of 0.59 or less

    -9 (4%) hospitals with a reliability score between 0.60 - 0.69 or less

    Panel Member #5: BetaBinomial. Mean reliability 0.9, nearly all >0.6.

    Panel Member #7: See #8 above

    Panel Member #9: The reliability testing, based on hospitals with at least 200 births, are not sufficient assess measure for fewer births, but statistics tells us that measures for small hospitals will have less. This problem can be solved by only using the measure for at least 200 births

    Panel Member #10: SNR based on the beta binomial model. Overall reliability is excellent (0.90). Looked at distribution of reliability by decile (presumably deciles here were based on actual reliability rating. This is not ideal. Should have instead grouped hospitals based on deciles of case volume, and then evaluated

reliability to determine what the volume threshold is for reliability to be equal to or greater than 0.70. This approach is suggested but certainly not required in the NQF white paper.)

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

    Submission document: Testing attachment, section 2b2.

    Panel Member #1: Developers supported the exclusion criteria with strong face validity concerning the measure's intention, which is to assess healthy, full term neonates. I assume 'non-malformation exclusion' in Table 7 refers to 'pre-existing conditions exclusion' mentioned in 2b2.3?

    Panel Member #2: Huge tranches of exclusions (and additions) since last submission. all seem reasonable but I am not an informed clinical reviewer

    Panel Member #3: No concerns

    Panel Member #4: Concerned in that cases with "social indications" are excluded for the following reasons:

    [1] Cases include various vulnerable populations (e.g. homelessness, extreme poverty) and should be included unless adequately explained. However, no explanation provided.

    [2] Such cases are not stated as excluded from the denominator (MIF, S.7), but are noted as numerator exclusions (MIF, S.5, step "e"). Essentially a rule of measurement is cases in the denominator must be at risk for the numerator event. In this measures, that's not the case. Leaving these cases in the denominator, but given they can't be counted in the numerator basically artificially inflates the denominator. In other words, it waters down the rate. This is problematic in a number of ways, such as hospitals with large numbers of cases with a "social indication" are predisposed to lower rates of the adverse event.

    Panel Member #7: None

    Panel Member #8: Exclusions included premature babies, low birth weight babies, congenital malformations, preexisting conditions and maternal drug use. These seem rational to me because the measure is expected to assess unexpected complications. However, the overall frequency of exclusions is 14.7%, which seems high to me.

    Panel Member #9: No concerns

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

    Submission document: Testing attachment, section 2b4

    Panel Member #1: No concerns. There seems to be a wide variation of performance among hospitals.

    Panel Member #2: None noted. 20% worse than state mean and 38% better

    Panel Member #3: No concerns.

    Panel Member #4: No concerns as there is a reasonable degree of variation as cited below: "…19.7% … rated as statistically significantly higher (worse) and …38.0% were identified as statistically significantly lower (better)…" [p22]

    Panel Member #7: It is difficult to assess the magnitude of between hospital differences vs. error variance at all ranges of the distribution.

    Panel Member #8: Hospitals were compared to CA state means. Differences were detected and I have no concerns.

    Panel Member #9: No concerns

Panel Member #10: Because this measure is not risk-adjusted, it does not take into account differences in hospital performance that may be due to differences in case mix. The measure developers (MD) have not empirically justified the decision not to risk adjust. There is empirical evidence that maternal characteristics, such as age, BMI, parity, history or prior CS, placental abnormality etc.) is associated with severe newborn morbidity. Virtually all other outcome measures used to evaluate hospital quality (ACS NSQIP, CMS measures, STS measures) adjust for patient factors. The absence of risk adjustment is a major limitation of this measure.

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

    Panel Member #1: NA
    Panel Member #2: NA
    Panel Member #3: Not applicable
    Panel Member #4: NA
    Panel Member #7: N/A
    Panel Member #8: Not applicable
    Panel Member #9: No concerns

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

    Submission document: Testing attachment, section 2b6

    Panel Member #1: Overall, missing data were rare and well handled.
    However, for cost analyses, 30 of the 238 California hospitals were not included because of missing charges data. Although this is not a huge percentage of missing data (13%), some comparison between hospitals with or without charge data would have enabled to assess the validity of the cost analyses presented.

    Panel Member #2: No concern…rates very low

    Panel Member #3: No concerns

    Panel Member #4: No concerns
    Panel Member #7: None

    Panel Member #8: I feel like checking the birth certificate to check for missing gestational age as well as including 2 sources of gestational age and adding the length of stay sufficiently addresses missing data and I have no concerns.

    Panel Member #9: 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

    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

    Panel Member #4: NA – not risk adjusted

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

    Panel Member #4: NA – not risk adjusted

Panel Member #10: No attempt has been made to empirically evaluate the inclusion of social factors on hospital performance because this model is not risk adjusted. It is not sufficient to simply state that inclusion of social risk factors could mask below-average care of vulnerable populations. The MD need to examine whether (1) social risk factors are associated with severe newborn morbidity and (2) whether including or not including social risk factor(s) in a risk-adjustment model has a substantial effect on hospital performance.

16d.**Risk adjustment summary:**

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

Panel Member #4: NA – not risk adjusted

16d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion?

☐ Yes ☐ No

Panel Member #4: NA – not risk adjusted

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

Panel Member #4: NA – not risk adjusted

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

☒ Yes ☒ No

Panel Member #4: NA – not risk adjusted

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

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

Panel Member #10: Because this measure is not risk-adjusted, it does not take into account differences in hospital performance that may be due to differences in case mix. The measure developers (MD) have not empirically justified the decision not to risk adjust. There is empirical evidence that maternal characteristics, such as age, BMI, parity, history or prior CS, placental abnormality etc.) is associated with severe newborn morbidity. Virtually all other outcome measures used to evaluate hospital quality (ACS NSQIP, CMS measures, STS measures) adjust for patient factors. The absence of risk adjustment is a major limitation of this measure.

Panel Member #9: The application indicates that there is risk assessment "Stratification by Single low risk strata" but doesn't really explain this.

Panel Member #8: I feel like this measure was developed when NQF had a different policy on risk stratification – and now that policy has changed. Because of this, I was hoping that the developer/steward would make more of an effort to test some risk adjustment strategies and compare them to the stratification techniques used. I would have liked to have seen a risk adjustment model include variables such as race, ethnicity and insurance status evaluated for inclusion in a risk model to assesses what the impact might be.

Panel Member #7: The argument for stratification by AAP NICU level vs. risk adjustment apparently based on the inaccurate statement "The National Quality Forum prefers that measures are not risk adjusted" for SES. It is not clear that the stratification analysis justifies the decision not to risk adjust results.

Panel Member #4: NA – not risk adjusted

Panel Member #1: Developers selected to not risk-adjust due to their use of exclusion criteria that are assumed to ensure a similar population between hospitals of healthy newborns not expected to have complications. Although 'risk-stratification' was marked, semantically I don't think this can be viewed as a risk-stratification approach since there is only one strata assessed.

Reasonable justification was provided for not risk-adjusting for social risk-factors, to avoid masking disparities that are not out of the hospitals' control concerning measure 0716. To some extent, this was also supported by analysis that confirmed similar distribution of hospital rates of unexpected complications by Levels of neonatal care units.

Panel Member #2: Developer argues that social risk factors are accounted for by exclusions (preterm growth, small-for-date infants, maternal substance abuse, etc), and therefore no need for social risk adjustment. It seems that other risk factors associated with SES (nutritional status, e.g.) are not adjusted among included cases. A concern but I'm not sure this should stop it from going forward.

Panel Member #3: The measure is constructed to create a homogenous group of patients; they did explore whether there are differences on the measure based on the hospital's NICU level and did not find any notable differences.

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 2b1.2

    Panel Member #1: Methods are appropriate at both data elements (patient level) and measure score level.
    Although not labeled as such, data element validity was reported in testing a & c sections.
    Length of stay (LOS) and hospital costs were compared between newborns with or without unexpected complications and supported the assumption that those with complications were expected to have higher LOS and costs. Note a typo on numbering of Table 3 ("34").
    Additionally, 'testing (c)' section reports on the stability of inclusion rates in the denominator over time when ICD codes transitioned from ICD-9 to ICD-10, as well as the stability of unexpected complication rates during 2014-2017. To some extent this also supports data element validity.
    Table 6 notes 'rate per 1,000. Is this an error, as the rate is reported in percent?
    At the score (hospital) level, a moderate and significant correlation around 0.4 was observed between average rates of unexpected newborn complications and average hospital LOS & costs. Since these measures, although related, are conceptually different, a moderate correlation is expected and in my view sufficiently supports empirical validity at the score level.
    A higher correlation of 0.64 was observed for a subset of cases including 49 hospitals between hospital rates of unexpected newborn complications and hospital rates of NICU admission (testing d).
    *Note that on Table 5, I assume that as for Table 4, there were a number of hospitals not included in the cost analyses. Please clarify.
    A Minor issue: Note that 'Systematic assessment of face validity' was checked but not reported.

    Panel Member #2: Association of measure with cost of care and length of stay…also with NICU admissions

    Panel Member #3: Hypothesized that hospital rates of Unexpected Newborn Complications would be correlated with NICU admission, newborn LOS, and cost; did not test this same hypothesis for Integrated Delivery Systems or Populations.

    Panel Member #4: Concern follows: Use of "LOS" seems illogical in some ways for construct validity testing for this particular measure. Given LOS is a numerator event, it seems the measure steward is not really demonstrating anything to essentially demonstrate that LOS is correlated to an adverse event where that numerator includes LOS. In evaluating this analysis it would help to know the portion of adverse events that are LOS. For example, if the numerator only contained 1% of cases due to LOS then this analysis would be more meaningful.

Otherwise, face validity is noted. However in 2b1.2 the process and structure of this testing is not explained. So unable to comment on how this method was employed. Further, in 2b1.3 there's no discussion of face validity findings.

'"We examined several options for construct validity testing. NICU admission as a comparator was not our first choice because of data indicating large variation (40 -fold!) among hospitals for NICU admission for term infants (Schulman J et al, 2018) that appeared to be primarily related to bed availability. Hospital cost (adjusted from charges) has been used as a reliable marker of morbidity including recent studies with Severer Maternal Morbidity (ChenHY et al, 2018). Similarly, length of stay can be used an independent marker of degree of illness. (Snowden CP et al, 2013). We tested Unexpected Newborn Complications measure against each of these three alternative measures for construct validity." [p14]

Panel Member #7: Adequate

Panel Member #8: Empirical validity testing of measure score was assessed looking at: 1) an assessment of the measure and length of stay and hospital costs (patient level data); 2) a Perarson correlation between the measure and average length of stay and costs (hospital level data); 3) comparing the measure in the ICD-9 data period and in the ICD-10 data period; and 4) a Perarson correlation between the measure and NICU admission rate of (via chart review). I have no concerns about these methods.

Panel Member #9: The application has "Systematic assessment of face validity" checked, but doesn't offer any evidence of this. Empirically, the developers assess this measure is correlated with NICU admission, newborn length of stay, and newborn hospital costs, as would be expected.

Panel Member #10: Examined the correlation between rate of unexpected complications with:

- Cost (correlation coefficient [r] = 0.4)
- LOS (r = 0.37)
- NICU admission (r = 0.64)

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

Submission document: Testing attachment, section 2b1.3, 2b1.4

Panel Member #1: Overall these analyses sufficiently support the empirical validity on 0716 score level.

Panel Member #2: Pearson corrs range 0.37-0.64, p values significant (<0.01) in Wilcoxin 2-sample t-tests comparing those with and without unexpected complications

Panel Member #3: Hypothesized that hospital rates of Unexpected Newborn Complications would be correlated with NICU admission, newborn LOS, and cost; did not test this same hypothesis for Integrated Delivery Systems or Populations.

Panel Member #4: While the tests performed are arguable (as discussed in Q21 above), the results of the construct validity testing are modest in general.

"Testing a. The mean and median length of stay and newborn hospital cost were markedly higher among term babies with unexpected complications compared to those without complications…. Wilcoxin two-sample test… $P < 0.01$, Table 3 & Table 4" [p15]

Testing b. …Table 5. Pearson Correlation coefficient:

-LOS: 0.41

-hospital cost: 0.37" [p15]

"Testing c. … inclusion rate was stable from the ICD-9 to the ICD-10" [p15]

"Testing d. The correlation coefficient of the rate of unexpected newborn complication and NICU admission was 0.64" [p16]

Panel Member #7: Unexpected results appear to support the relationship between the newborn complication rate and LOS and hospital cost validation variables.

Panel Member #8: No concerns

Panel Member #9: The application has "Systematic assessment of face validity" checked, but doesn't offer any evidence of this.

Empirically, the developers assess this measure is correlated with NICU admission, newborn length of stay, and newborn hospital costs, as would be expected.

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)

Panel Member #4: Response to Q21 above: Concern follows: Use of "LOS" seems illogical in some ways for construct validity testing for this particular measure. Given LOS is a numerator event, it seems the measure steward is not really demonstrating anything to essentially demonstrate that LOS is correlated to an adverse event where that numerator includes LOS. In evaluating this analysis it would help to know the portion of adverse events that are LOS. For example, if the numerator only contained 1% of cases due to LOS then this analysis would be more meaningful.

Otherwise, face validity is noted. However in 2b1.2 the process and structure of this testing is not explained. So unable to comment on how this method was employed. Further, in 2b1.3 there's no discussion of face validity findings.

Panel Member #9: The results show that this measure is correlated with NICU admission, newborn length of stay, and newborn hospital costs, as would be expected.

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.**

Panel Member #1: Validity was supported both at the data element and score levels, with minimal threats to validity.

Panel Member #3: If measure is going to be specified for Integrated Delivery Systems or Populations: Regional or State, validity testing should be provided for that level of analysis.

Panel Member #4: Response to Q21 above: Concern follows: Use of "LOS" seems illogical in some ways for construct validity testing for this particular measure. Given LOS is a numerator event, it seems the measure steward is not really demonstrating anything to essentially demonstrate that LOS is correlated to an adverse event where that numerator includes LOS. In evaluating this analysis it would help to know the portion of adverse events that are LOS. For example, if the numerator only contained 1% of cases due to LOS then this analysis would be more meaningful.

Otherwise, face validity is noted. However in 2b1.2 the process and structure of this testing is not explained. So unable to comment on how this method was employed. Further, in 2b1.3 there's no discussion of face validity findings.

Response to Q22 above: While the tests performed are arguable (as discussed in Q21 above), the results of the construct validity testing are modest in general.

Panel Member #5: Comparison of numerator with newborn LOS, cost, NICU admission, ICD9 vs ICD10. Rationale and attempted empirical justification for no risk adjustment provided. The clinical expertise of this group exceeds mine. I do wonder if variables such as BMI or attendance at prenatal visits would explain some variation in outcomes.

Panel Member #7: The choice of validation variables and results from analyses support the validity of this measure.

Panel Member #8: No concerns

Panel Member #9: The results show that this measure is correlated with NICU admission, newborn length of stay, and newborn hospital costs, as would be expected.

Despite being promised, face validity testing does not seem to have been conducted, but that doesn't influence the results.

Panel Member #10: Because this measure is not risk-adjusted, it does not take into account differences in hospital performance that may be due to differences in case mix. The measure developers (MD) have not empirically justified the decision not to risk adjust. There is empirical evidence that maternal characteristics, such as age, BMI, parity, history or prior CS, placental abnormality etc.) is associated with severe newborn morbidity. Virtually all other outcome measures used to evaluate hospital quality (ACS NSQIP, CMS measures, STS measures) adjust for patient factors. The absence of risk adjustment is a major limitation of this measure.

It is not sufficient to demonstrate that severe newborn morbidity is correlated with the rate of NICU admissions or other similar measures. This type of validity evaluation sets an extremely low bar. A priori, there is no reason to believe that some of the differences between hospital performance based on this measure are due to differences in case mix as opposed to differences in actual hospital quality. The measure developers need to demonstrate empirically that this is not the case.

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.**

Panel Member #5: The specifications include major and minor outcomes in the numerator. I may have missed this, but how are such dealt with in the numerator (two measures or combined?). (Curious to learn Larry's take on this given his work in this area.)

Developer Submission

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

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

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

NQF_evidence_attachment.UNC-04.03.2020.docx

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

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

Yes

1a. Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0716

Measure Title: Unexpected Newborn Complications in Term Infants

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: Click here to enter a date

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

Outcome

☒ Outcome: Quality of obstetric and neonatal care at the hospital level

☐ 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

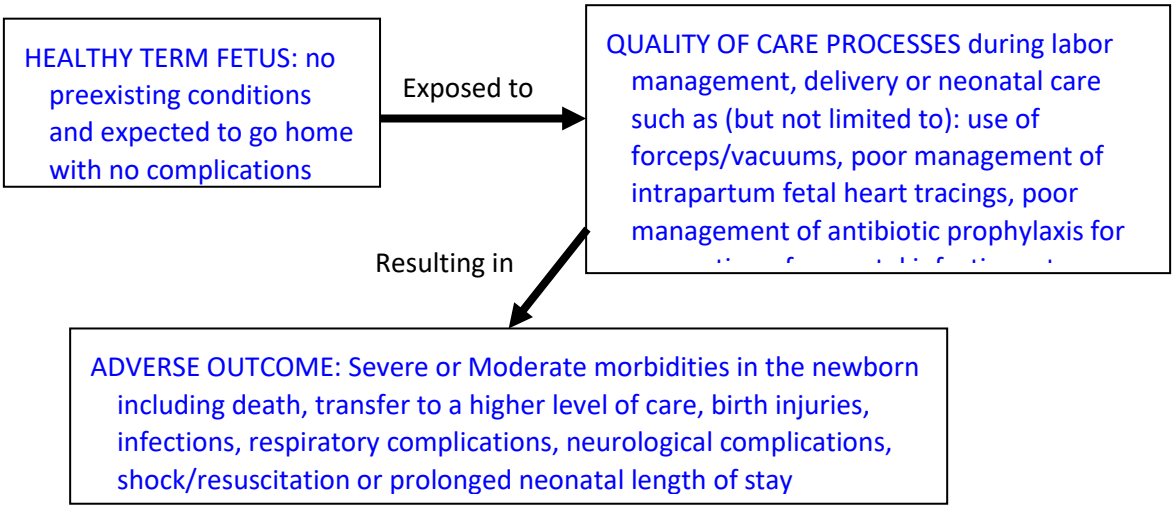
☐ Composite: Click here to name what is being measured

1a.2 LOGIC MODEL 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.

In the panel of maternal and child health outcomes, this measure is the one that evaluates the largest population of live births; healthy term newborns. This metric also serves as a balancing measure for other

maternal and neonatal procedures and measures (i.e.) if you increase (or decrease) one does that affect baby outcomes? For example, many modern obstetric practices (such as the use of inductions, vacuums, forceps and cesarean deliveries) are done in the name of improving baby outcomes without having a proper measure to document that. In fact, many of these interventions may negatively impact newborn health in some settings. Hence the importance of following a global measure of newborn outcomes for every hospital and not just at the population level.

Multiple care processes can influence a deterioration in a newborn's health status during labor management, delivery or neonatal care resulting in unexpected severe or moderate morbidities for the newborn with potential short or long term consequences.



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.

1a.3. SYSTEMATIC REVIEW(SR) OF THE EVIDENCE (for INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURES, INCLUDING THOSE THAT ARE INSTRUMENT-BASED) If the evidence is not based on a systematic review go to section 1a.4) If you wish to include more than one systematic review, add additional tables.

What is the source of the systematic review of the body of evidence that supports the performance measure? A systematic review is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but

separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data. (IOM)

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	
Provide all other grades and definitions from the evidence grading system	
Grade assigned to the **recommendation** with definition of the grade	
Provide all other grades and definitions from the recommendation grading system	
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	
Estimates of benefit and consistency across studies	
What harms were identified?	
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	

_____

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.

Five studies in the last 18 months have used Unexpected Newborn Complications as either a key outcome or important balancing measure during studies focused on improving obstetric practice and offer comparisons to other simultaneously collected neonatal outcome measures.

(1) Shields (2018) implemented a protocol to standardize the response to Category II Fetal Heart Rate patterns in 6 hospitals. Their new protocol showed improved outcomes when compared to baseline: 5-minute Apgar scores <7 were reduced by 24.6% and Severe Unexpected Newborn Complications scores were reduced by 26.6% accompanied by a slight decrease in the cesarean rate (19.8% to 18.3%).

(2) Xu (2019) examined state-wide California data for neonatal outcomes following attempted vaginal birth after prior cesarean delivery. After adjustment for patient risk factors, those delivered at hospitals with above-the-median utilization and success rates of trial of labor had a higher risk for uterine rupture (adjusted risk ratio, 2.74, P < .001), and, using the CMQCC recommend UNC subsets, severe newborn respiratory complications (adjusted risk ratio, 1.46, P < .001), and severe newborn neurological complications/trauma (adjusted risk ratio, 2.48, P < .001), but they had a lower risk for severe newborn infection (adjusted risk ratio, 0.80, P = .003) and overall Severe Unexpected Newborn Complications (adjusted risk ratio, 0.86, P < .001) as well as shorter length of stays (adjusted mean ratio, 0.948 for mothers and 0.924 for newborns, P < .001 for both).

(3) Kahwati (2019) reported on a large scale AHRQ study (43 hospitals) using Team Steps to help drive perinatal safety. Statistically significant decreases in indicators for obstetric trauma without instruments and primary cesarean delivery were observed. A statistically significant increase in neonatal birth trauma was observed, but the overall rate of Unexpected Newborn Complications was unchanged. They concluded that the program had a favorable impact on unit patient safety culture and processes, but short-term impact on maternal and neonatal adverse events was mixed.

(4) Main (2019) used Severe Unexpected Newborn Complications as a balancing measure for a large-scale quality improvement collaborative to reduce primary cesarean births (56 hospitals, 119,000 annual births). Among collaborative hospitals, the nulliparous, term, singleton, vertex (NTSV) cesarean delivery rate fell from 29.3% in 2015 to 25.0% in 2017 (2017 vs 2015 adjusted OR [aOR] 0.76, 95% CI 0.73-0.78). None of the safety measures (Severe Unexpected Newborn Complications, 5-minute Apgar Score <5, chorioamnionitis rate, transfusion rate, and 3rd or 4th degree laceration rate) showed any difference comparing 2017 to 2015. As a sensitivity analysis, the tercile of hospitals with the greatest decline in NTSV cesarean rates (31.2% to 20.6%, 2017 vs 2015 aOR 0.54, 95% CI 0.50-0.58) was examined to evaluate whether they had greater risk of poor maternal and neonatal outcomes. Again, no measure was statistically worse, and the Severe Unexpected Newborn Complications composite actually improved (3.2% to 2.2%, aOR 0.71, 95% CI 0.55-0.92).

(5) Kuhlmann-Capek (2020) on behalf of the NICHD MFMU Network reported an analysis examining the relationship between Severe Unexpected Newborn Complications and Umbilical artery base deficit (UABD) in nearly 10,000 term infants. There was a significant association between UABD and both moderate and severe complications, even after adjustment for patient characteristics and cesarean delivery. The association was even stronger for severe than moderate and very predictive for the higher quartiles of UABD. For UABD quartile 3, the aOR was 4.24 and for UABD quartile 4, the OR was 32.01.

These studies support the correlation between Total Unexpected Newborn Complications and Severe Unexpected Newborn Complications with other commonly used neonatal outcome metrics (such as Apgar scores, birth injuries, and umbilical artery base deficit). CMQCC findings presented in Tables 3 and 4 in Section 2b1.3, Validity Testing of the Measure, illustrate a good correlation between Unexpected Newborn Complications and neonatal costs (as good marker of morbidity) and neonatal LOS (both $p < 0.01$). The findings presented here are important as umbilical blood gases, birth injuries and even NICU admissions are much more difficult to routinely collect and typically represent a more narrow range of concerning neonatal conditions than the more broadly configured Unexpected Newborn Complications. Moving beyond correlations with other neonatal outcomes, several of the studies described here showed improvements in Unexpected Newborn Complications following large-scale quality improvement/safety initiatives. This illustrates that actionability of Unexpected Newborn Complications.

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

1. Internal CMQCC research using CA, WA and OR data sets (some of which is presented in section
2. PubMed search April 1, 2020, see below, section 1a.4.3

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

Recent UNC Literature:

(1) Shields LE, Wiesner S, Klein C, Pelletreau B, Hedriana HL. A Standardized Approach for Category II Fetal Heart Rate with Significant Decelerations: Maternal and Neonatal Outcomes. Am J Perinatol. 2018 Dec;35(14):1405-1410.

(2) Xu X, Lee HC, Lin H, Lundsberg LS, Campbell KH, Lipkind HS, Pettker CM, Illuzzi JL. Hospital variation in utilization and success of trial of labor after a prior cesarean. Am J Obstet Gynecol. 2019 Jan;220(1):98.e1-98.e14.

(3) Kahwati LC, Sorensen AV, Teixeira-Poit S, Jacobs S, Sommerness SA, Miller KK, Pleasants E, Clare HM, Hirt CL, Davis SE, Ivester T, Caldwell D, Muri JH, Mistry KB. Impact of the Agency for Healthcare Research and Quality's Safety Program for Perinatal Care. Jt Comm J Qual Patient Saf. 2019 Apr;45(4):231-240.

(4) Main EK, Chang SC, Cape V, Sakowski C, Smith H, Vasher J. Safety Assessment of a Large-Scale Improvement Collaborative to Reduce Nulliparous Cesarean Delivery Rates. Obstet Gynecol. 2019 Apr;133(4):613-623.

(5) Kuhlmann-Capek MJ, for the Eunice Kennedy Shriver National Institute of Child Health and Human Development Maternal-Fetal Medicine Units Network, Bethesda, MD. Relationship between "Unexpected Complications in Term Newborns" perinatal quality measure and umbilical artery base deficit. Am J Obstet Gynecol 2020;222:S44-45

1b. Performance Gap

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

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

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

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

The most important childbirth outcome for families is bringing home a healthy baby. While there have been measures developed to assess clinical practices and outcomes in preterm infants, there are a lack of metrics that assess the health outcomes of term infants who represent over 90% of all births.

The Unexpected Complications in Term Newborns metric addresses this gap and measures adverse outcomes resulting in severe or moderate morbidity in otherwise healthy term infants without preexisting conditions. Importantly, this metric also serves as a balancing measure for other NQF endorsed maternal measures such as NTSV Cesarean rates, third and fourth degree lacerations, episiotomies and early elective delivery rates. The purpose of a balancing measure to is guard against any unanticipated or unintended consequences of quality improvement activities for these measures.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for maintenance of endorsement. 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.*

2017 data

Dates of data: 1st Jan 2017- 31st Dec 20117

Number of hospitals in California that had births in 2017: n=238

Number of hospitals in California that had eligible cases in the denominator: n=238

Total number of eligible births (included in the denominator): n=373,763

Total number of hospitals with over 200 cases in the denominator: 225

Total number of eligible births in hospitals with over 200 cases in the denominator: n=372,139

Distribution of the rate of unexpected newborn complications (225 hospitals with over 200 cases in the denominator)

Mean	Std Dev	Median	Q1	Q3	Min	Max
3.19	1.88	2.74	1.84	4.07	0.21	11.21

Distribution of the rate of unexpected newborn complications (225 hospitals with over 200 cases in the denominator), by deciles

Deciles	N of hospitals	Mean	Std Dev	Minimum	Maximum
1	21	1.02	0.37	0.21	1.37
2	24	1.52	0.08	1.38	1.65
3	22	1.86	0.10	1.71	2.03
4	22	2.19	0.09	2.07	2.32
5	24	2.53	0.13	2.33	2.74
6	23	2.98	0.15	2.75	3.22
7	22	3.49	0.21	3.25	3.85
8	23	4.09	0.12	3.90	4.3
9	22	4.75	0.28	4.37	5.34
10	22	7.52	1.64	5.39	11.21

TIME COURSE DATA FOR UNC in California (state-wide, including all hospitals):

rate per 1,000 births

2013 32.3

2014. 32.3

2015. 30.4

2016. 29.7

2017. 29.3

2018. 30.0

2019. 29.1

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.

Performance Data has been displayed in Question 1.b.2. The data has been produced and tested on over 4 years of administrative data in California. This measure is also in use in Oregon, Washington and by hospitals across the United States who are part of the National Perinatal Information Center´s Network of hospitals. In addition, this measure has been added to be one of the Joint Commission National Quality Measures since the version v2018B.

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.*

See tables below for 2017 data

Rate of unexpected newborn complications in 225 hospitals with over 200 cases in the denominator in 2017, by newborn and hospital characteristics

Characteristics	Numerator (N=11,547)	Denominator (N=372,139)	Unexpected newborn complication rate (%)
Neonatal sex			
Female	4,803	182,123	2.64
Male	6,744	190,016	3.55
Maternal racial/ethnic group			
White	3,283	102,851	3.19
Black	680	18,635	3.65
Asian	1,541	59,267	2.60
Hispanics	5,556	177,520	3.13
Others	108	3,271	3.30
Missing	379	10,595	3.58
Maternal insurance status			
Medi-Cal	5,862	171,975	3.41
Private insurance	5,408	185,879	2.91
Self-pay	277	14,285	1.94
AAP neonatal level of care			
Level I	1,205	39,763	3.03
Level II	2,623	79,525	3.30

Level III	6,279	205,996	3.05
Level IV	1,440	46,855	3.07
Geographic location of hospitals			
Non-urban	361	10,526	3.43
Urban	11,186	361,613	3.09
Hospital ownership			
University	624	14,521	4.30
County	374	8,473	4.41
District	1,033	30,253	3.41
Integrated Health System	1,607	67,963	2.36
Private non-profit	6,448	201,174	3.21
Private investor	1,461	49,755	2.94

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

Data is provided in 1b.4

2. Reliability and Validity—Scientific Acceptability of Measure Properties

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

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

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

Perinatal Health, Perinatal Health : Newborn Care

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

Safety : Complications

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

Children, Women

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

https://www.cmqcc.org/focus-areas/quality-metrics/unexpected-complications-term-newborns

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:** Unexpected_Newborn_Complications_Measure_Specifications-Jan5.xlsx

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

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

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

Not an instrument-based measure

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

Yes

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

We have enclosed all of the changes made to the measure to transition from ICD9 to ICD10; to keep up with annual changes in codes; to respond to learning from out over 300 hospital user group and to transition away from the use of the birth certificate. See the Summary of UNC changes tab in the excel spreadsheet in S.2b above.

After reviewing data from three states (CA, WA and OR) and consultation with our expert panel, we found that many babies with these two codes actually were quite healthy and had only transient concern. In addition, the CPAP code was being used often for bag and mask ventilation in the Delivery Room (not in accordance with Coding Clinic guidelines) to allow billing for a pediatrician to attend. Therefore, we decided to add a LOS modifier for both (ie for this code to be considered the LOS needed to exceed 2 days for a vaginal birth and 4 days for a Cesarean birth). These changes do lead to changes in the measure and sub-measures that will be discussed later. THESE ARE THE MOST IMPORTANT CHANGES OF ALL THE CHANGES MADE

1) Move this diagnosis code from Group 4A: Moderate Birth Trauma (Diagnosis Codes) to Group 4D: Moderate Birth Trauma with specific LOS requirement (Diagnosis Codes)

767.2 Fracture of clavicle due to birth trauma

2) Move this procedure code from Group 4C: Moderate Respiratory Complication (Procedure Codes) to Group 4G: Moderate Respiratory Complications with specific LOS requirement (Procedure Codes)

93.90 Non-invasive mechanical ventilation (e.g. CPAP)

Part 3:

We removed all of the codes for cerebral palsy as the diagnosis should not be made until much later in infancy (not in the immediate neonatal period) and on review there were only 1-2 cases out of over a million births. Chorioamnionitis affecting the newborn (associated with a prolonged neonatal LOS) was added because of its increasing incidence and to create parity with ICD-10.

1) Remove these 7 diagnosis codes from Group 3B: Severe Hypoxia Asphyxia (Diagnosis Codes)

343.0 Congenital diplegia

343.1 Congenital hemiplegia

343.2 Congenital quadriplegia

343.3 Congenital monoplegia

343.4 Infantile hemiplegia

343.8 Other specified infantile cerebral palsy

343.9 Infantile cerebral palsy, unspecified

2) Add this diagnosis code to Group 3J: Severe Septicemia (LOS > 4 DAYS): (Diagnosis Codes)

762.7 Chorioamnionitis affecting fetus or newborn

ICD 10 Changes

Part 1:

Following a review of our initial experience with UNC ICD-10 codes we discovered a much higher rate of exclusions for congenital malformations than under ICD-9. Case reviews found that the much-expanded ICD-10 anomaly list led to exclusions for minor conditions that did not affect the immediate neonatal health. Therefore, the list was reviewed by an expert panel of neonatologists who recommended removing the following codes from our exclusion list.

1) Remove the following 103 diagnosis codes from Group 2A: Congenital Malformations (includes disorders and syndromes)

E80.6 Other disorders of bilirubin metabolism

Q10.5 Congenital stenosis and stricture of lacrimal duct

Q16.0 Congenital absence of (ear) auricle

Q16.1 Congenital absence, atresia and stricture of auditory canal (external)

Q16.2 Absence of eustachian tube

Q16.3 Congenital malformation of ear ossicles

Q16.4 Other congenital malformations of middle ear

Q16.5 Congenital malformation of inner ear

Q16.9 Congenital malformation of ear causing impairment of hearing, unspecified

Q17.0 Accessory auricle

Q17.1 Macrotia

Q17.3 Other misshapen ear

Q17.4 Misplaced ear

Q17.5 Prominent ear

Q17.8 Other specified congenital malformations of ear

Q17.9 Congenital malformation of ear, unspecified

Q27.0 Congenital absence and hypoplasia of umbilical artery

Q38.0 Congenital malformations of lips, not elsewhere classified

Q38.1 Ankyloglossia

Q38.3 Other congenital malformations of tongue

Q38.4 Congenital malformations of salivary glands and ducts

Q52.10 Doubling of vagina, unspecified.

Q52.11 Transverse vaginal septum.

Q52.12 Longitudinal vaginal septum.

Q53.00 Ectopic testis, unspecified

Q53.01 Ectopic testis, unilateral

Q53.02 Ectopic testes, bilateral

Q53.10 Unspecified undescended testicle, unilateral

Q53.11 Abdominal testis, unilateral

Q53.12 Ectopic perineal testis, unilateral
Q53.20 Undescended testicle, unspecified, bilateral
Q53.21 Abdominal testis, bilateral
Q53.22 Ectopic perineal testis, bilateral
Q53.9 Undescended testicle, unspecified
Q54.0 Hypospadias, balanic
Q54.1 Hypospadias, penile
Q54.2 Hypospadias, penoscrotal
Q54.3 Hypospadias, perineal
Q54.4 Congenital chordee
Q54.8 Other hypospadias
Q54.9 Hypospadias, unspecified
Q55.0 Absence and aplasia of testis
Q55.1 Hypoplasia of testis and scrotum
Q55.20 Unspecified congenital malformations of testis and scrotum
Q55.21 Polyorchism
Q55.22 Retractile testis
Q55.23 Scrotal transposition
Q55.29 Other congenital malformations of testis and scrotum
Q55.3 Atresia of vas deferens
Q55.4 Other congenital malformations of vas deferens, epididymis, seminal vesicles and prostate
Q55.5 Congenital absence and aplasia of penis
Q55.61 Curvature of penis (lateral)
Q55.62 Hypoplasia of penis
Q55.63 Congenital torsion of penis
Q55.64 Hidden penis
Q55.69 Other congenital malformation of penis
Q55.7 Congenital vasocutaneous fistula
Q55.8 Other specified congenital malformations of male genital organs
Q55.9 Congenital malformation of male genital organ, unspecified
Q69.0 Accessory finger(s)
Q69.1 Accessory thumb(s)
Q69.2 Accessory toe(s)
Q69.9 Polydactyly, unspecified
Q70.00 Fused fingers, unspecified hand
Q70.01 Fused fingers, right hand
Q70.02 Fused fingers, left hand
Q70.03 Fused fingers, bilateral
Q70.10 Webbed fingers, unspecified hand

Q70.11 Webbed fingers, right hand

Q70.12 Webbed fingers, left hand

Q70.13 Webbed fingers, bilateral

Q70.20 Fused toes, unspecified foot

Q70.21 Fused toes, right foot

Q70.22 Fused toes, left foot

Q70.23 Fused toes, bilateral

Q70.30 Webbed toes, unspecified foot

Q70.31 Webbed toes, right foot

Q70.32 Webbed toes, left foot

Q70.33 Webbed toes, bilateral

Q70.4 Polysyndactyly, unspecified

Q70.9 Syndactyly, unspecified

Q82.1 Xeroderma pigmentosum

Q82.2 Mastocytosis

Q82.3 Incontinentia pigmenti

Q82.4 Ectodermal dysplasia (anhidrotic)

Q82.5 Congenital non-neoplastic nevus

Q82.8 Other specified congenital malformations of skin

Q82.9 Congenital malformation of skin, unspecified

Q83.0 Congenital absence of breast with absent nipple

Q83.1 Accessory breast

Q83.2 Absent nipple

Q83.3 Accessory nipple

Q83.8 Other congenital malformations of breast

Q83.9 Congenital malformation of breast, unspecified

Q84.0 Congenital alopecia

Q84.1 Congenital morphological disturbances of hair, not elsewhere classified

Q84.2 Other congenital malformations of hair

Q84.3 Anonychia

Q84.4 Congenital leukonychia

Q84.5 Enlarged and hypertrophic nails

Q84.6 Other congenital malformations of nails

Q84.8 Other specified congenital malformations of integument

Q84.9 Congenital malformation of integument, unspecified

Part 2:

We also noted a few overlooked codes that were missing from the initial set and two codes that were placed in the wrong category.

1) Add these two diagnosis codes to Group 3C: Severe Shock and Resuscitation (Diagnosis Codes)

I46.9 Cardiac arrest, cause unspecified

P29.0 Neonatal cardiac failure

2) Add this diagnosis code to Group 4D: Moderate Birth Trauma with LOS

P15.4 Birth injury to face

3) Move these two diagnosis codes from Group 3F: Severe Neurological Complications (Diagnosis Codes) to Group 3C: Severe Shock and Resuscitation (Diagnosis Codes)

P29.4 Transient myocardial ischemia in newborn

P29.81

Cardiac arrest of newborn

Part 3:

We noted that the ICD-10 codes that began: "Newborn (suspected to be) affected by…" were too non-specific and generally not associated with significant morbidity when used alone. Therefore, we decided to remove them.

1) Remove the two diagnosis codes from Group 2B: Other Fetal Placental Conditions (Diagnosis Codes)

P02.20 Newborn (suspected to be) affected by unspecified morphological and functional abnormalities of

P02.29 Newborn (suspected to be) affected by other morphological and functional abnormalities of place

2) Remove the six diagnosis codes from Group 2C: Maternal Drug Use (Diagnosis Codes)

P04.1 Newborn (suspected to be) affected by other maternal medication

P04.2 Newborn (suspected to be) affected by maternal use of tobacco

P04.5 Newborn (suspected to be) affected by maternal use of nutritional chemical substances

P04.6 Newborn (suspected to be) affected by maternal exposure to environmental chemical substances

P04.8 Newborn (suspected to be) affected by other maternal noxious substances

P04.9 Newborn (suspected to be) affected by maternal noxious substance, unspecified

3) Remove the following eight diagnosis codes from Group 4D: Moderate Birth Trauma with specific Los Requirement (>4D CS or >2D Vaginal) Diagnosis Codes

P02.5 Newborn (suspected to be) affected by other compression of umbilical cord

P03.5 Newborn (suspected to be) affected by precipitate delivery

P03.6 Newborn (suspected to be) affected by abnormal uterine contractions

P03.810 Newborn (suspected to be) affected by abnormality in fetal (intrauterine) heart rate or rhythm

P03.811 Newborn (suspected to be) affected by abnormality in fetal (intrauterine) heart rate or rhythm

P03.819 Newborn (suspected to be) affected by abnormality in fetal (intrauterine) heart rate or rhythm

P03.89 Newborn (suspected to be) affected by other specified complications of labor and delivery

P03.9 Newborn (suspected to be) affected by complication of labor and delivery, unspecified

Part 4:

As noted for ICD-9, after reviewing data from three states (CA, WA and OR) and consultation with our expert panel, we found that many babies with these two codes actually were quite healthy and had only transient concern. In addition, the CPAP code was being used often for bag and mask ventilation in the Delivery Room (not in accordance with Coding Clinic guidelines) to allow billing for a pediatrician to attend. Therefore, we decided to add a LOS modifier for both (ie for this code to be considered the LOS needed to exceed 2 days for a vaginal birth and 4 days for a Cesarean birth). These changes do lead to changes in the measure and sub-measures that will be discussed later.

1) Move this diagnosis code from Group 4A: Moderate Birth Trauma (Diagnosis Codes) to Group 4D: Moderate Birth Trauma with specific LOS requirement (Diagnosis Codes)

P13.4 Fracture of clavicle due to birth injury

2) Move this procedure code from Group 4C: Moderate Respiratory Complication (Procedure Codes) to Group 4G: Moderate Respiratory Complications with specific LOS requirement (Procedure Codes)

5A09357 Assistance with Respiratory Ventilation, Less than 24 Consecutive Hours, Continuous Positive Airway Pressure

Part 5:

We removed all of the codes for cerebral palsy as the diagnosis should not be made until much later in infancy (not in the immediate neonatal period) and on review there were only 1-2 cases out of over a million births.

1) Remove these 7 diagnosis codes from Group 3B: Severe Hypoxia Asphyxia (Diagnosis Codes)

G80.0 Spastic quadriplegic cerebral palsy

G80.1 Spastic diplegic cerebral palsy

G80.2 Spastic hemiplegic cerebral palsy

G80.3 Athetoid cerebral palsy

G80.4 Ataxic cerebral palsy

G80.8 Other cerebral palsy

G80.9 Cerebral palsy, unspecified

2) Remove this diagnosis code from Group 4H: Moderate Infection Complications with Specific Los Requirement (> 4D CS OR > 2D Vaginal) Diagnosis Codes. This code is already in the Group Severe Sepsis with LOS >4d which identifies the severe cases, this picks up cases that are vaginal births and have LOS of 3or 4 days whcihc suggests they were getting an evaluation but not proven septic.

P02.7 Newborn (suspected to be) affected by chorioamnionitis

Part 6:

We removed the following procedure codes from the initial ICD-10 set after feedback from our neonatology panel and the extreme rarity of their use in the neonatal period.

1) Remove these 8 procedure codes from Group 3G: Severe Shock and Resuscitation (Procedure Codes).

0DH632Z Insertion of Monitoring Device into Stomach, Percutaneous Approach

0DH633Z Insertion of Infusion Device into Stomach, Percutaneous Approach

0DH63DZ Insertion of Intraluminal Device into Stomach, Percutaneous Approach

0DH63MZ Insertion of Stimulator Lead into Stomach, Percutaneous Approach

0DH642Z Insertion of Monitoring Device into Stomach, Percutaneous Endoscopic Approach

0DH643Z Insertion of Infusion Device into Stomach, Percutaneous Endoscopic Approach

0DH64DZ Insertion of Intraluminal Device into Stomach, Percutaneous Endoscopic Approach

0DH64MZ Insertion of Stimulator Lead into Stomach, Percutaneous Endoscopic Approach

2) Remove these 13 procedure codes from Group 3I: Severe Neurological Complications (Procedure Codes).

0D16074 Bypass Stomach to Cutaneous with Autologous Tissue Substitute, Open Approach

0D160J4 Bypass Stomach to Cutaneous with Synthetic Substitute, Open Approach

0D160K4 Bypass Stomach to Cutaneous with Nonautologous Tissue Substitute, Open Approach

0D160Z4 Bypass Stomach to Cutaneous, Open Approach

0D163J4 Bypass Stomach to Cutaneous with Synthetic Substitute, Percutaneous Approach

0D16474 Bypass Stomach to Cutaneous with Autologous Tissue Substitute, Percutaneous Endoscopic Approach

0D164J4 Bypass Stomach to Cutaneous with Synthetic Substitute, Percutaneous Endoscopic Approach

0D164K4 Bypass Stomach to Cutaneous with Nonautologous Tissue Substitute, Percutaneous Endoscopic Appro

0D164Z4 "Bypass Stomach to Cutaneous, Percutaneous Endoscopic Approach

0D16874 Bypass Stomach to Cutaneous with Autologous Tissue Substitute, Via Natural or Artificial Openin

0D168J4 Bypass Stomach to Cutaneous with Synthetic Substitute, Via Natural or Artificial Opening Endosc

0D168K4 Bypass Stomach to Cutaneous with Nonautologous Tissue Substitute, Via Natural or Artificial Ope

0D168Z4 "Bypass Stomach to Cutaneous, Via Natural or Artificial Opening Endoscopic

Part 7:

New ICD-10 codes have been released to take effect in October 2017.

1) Add these new diagnosis codes to Group 2A: Congenital Malformations (Exclusions). They represent expansions of four prior codes (E78.0, Q25.2, Q25.4 and Q66.2).

E78.00 Pure hypercholesterolemia, unspecified

E78.01 Familial hypercholesterolemia

Q25.21 Interruption of aortic arch

Q25.29 Other atresia of aorta

Q25.40 Congenital malformation of aorta unspecified

Q25.41 Absence and aplasia of aorta

Q25.42 Hypoplasia of aorta

Q25.43 Congenital aneurysm of aorta

Q25.44 Congenital dilation of aorta

Q25.45 Double aortic arch

Q25.46 Tortuous aortic arch

Q25.47 Right aortic arch

Q25.48 Anomalous origin of subclavian artery

Q25.49 Other congenital malformations of aortaQ66.21 Congenital metatarsus primus varus

Q66.22 Congenital metatarsus adductus

2) Add these new diagnosis codes to Group 2B: Other Fetal Placental Conditions (Exclusions). They represent expansions of the codes P29.3.

P29.30 Pulmonary hypertension of newborn

P29.38 Other persistent fetal circulation

3) Add these new procedure codes to Group 3G: Severe Shock and Resuscitation. They represent expansions of three prior codes (03HY0x, 03HY3x, and 03HY4x).

03HY0YZ Insertion of Other Device into Upper Artery, Open Approach

03HY3YZ Insertion of Other Device into Upper Artery, Percutaneous Approach

03HY4YZ Insertion of Other Device into Upper Artery, Percutaneous Endoscopic Approach

Impact of Fractured Clavicle and CPAP on Moderate UNC rates

As noted above, we extensively studied the codes for Fractured clavicle and Continuous Positive Airway Pressure (CPAP) and determined that they were over-valued in the sense that the large majority of the infants had minimal sequelae. This does impact Moderate UNC rates but not on Severe UNC (these codes were captured in Moderate Respiratory and Moderate Birth Trauma categories). The scale of difference in ICD-10 with the revised codes is shown below with the majority of the change driven by the CPAP code (in the Respiratory category):

Category	Original Code Set	Revised Code Set
Total UNC	3.76%	3.16%
Moderate Respiratory	1.32%	0.86%
Moderate Birth Trauma	0.28%	0.14%

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

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

Numerator: The numerator is divided into two categories: Severe complications and moderate complications.

Severe complications include neonatal death, transfer to another hospital for higher level of care, , severe birth injuries such as intracranial hemorrhage or nerve injury, neurologic damage, severe respiratory and infectious complications such as sepsis. Parents of such babies may often worry about short or long term infant outcomes.

Moderate complications include diagnoses or procedures that raise concern but at a lower level than the list for severe (e.g. use of CPAP or bone fracture). For inclusion in the numerator, most require an infant length of stay that exceeds that of the mother, validating that these are indeed significant complications. Examples include less severe respiratory complications (e.g. Transient Tachypnea of the Newborn), or infections with a longer length of stay not including sepsis. As a "safety net" to capture cases who were under-coded, the numerator also includes infants who have a prolonged length of stay of over 5 days to capture the "seemingly normal" infants with neither any form of jaundice nor a social reason for staying in the hospital (e.g. family disruption or adoption).

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

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

In the full term neonatal population that excluded premature infants, low birth weight babies, infants with congenital malformations, fetuses with pre-existing conditions such as IUGR and babies exposed to maternal drug use, babies were selected for inclusion in the numerator in a hierarchical manner as follows:

PART A: Severe Complications: Identify and include the following in a hierarchical manner:

a) Neonatal Deaths (Use patient discharge diagnosis data, specifically the disposition code for death)

b) Neonatal Transfers (Use patient discharge diagnosis data, specifically the disposition code for transfer to a higher level of care)

c) Severe Morbidities: (Use patient discharge diagnosis data, examining both primary and other diagnosis and procedure fields for ICD-10 Codes defining an array of specific severe complications. Please refer to Tables

11.36 thru 11.45 (Appendix 3, Groups 3A through 3I) with the specific ICD10 codes and descriptors listed in excel document in S.2b above and on our website.

d) Sepsis with a neonatal Length of Stay that exceeds 4 days (Use patient discharge diagnosis data, examining both primary and other diagnosis fields for the specific ICD-9 code defining sepsis. Note that neonatal stay is defined as the date of discharge minus the date of birth).

The neonates identified in Part A make up the "Severe Complications" component of the numerator.

In the remaining infants (those without severe morbidities), identify and include the following

PART B: Moderate Complications: Identify and include the following in a hierarchical manner:

a) Moderate complications not requiring a specific length of stay: Identify babies with moderate complications that do not require a specific length of stay for inclusion (Use Patient discharge Diagnosis data, examining both primary and other diagnosis and procedure fields for ICD-10 codes identifying specific moderate complications (see Table 11.46 thru Table 11.53 for the specific ICD10 codes and descriptors listed in excel document in S.2b above and on our website

b) Specific Prolonged neonatal length of Stay stratified by method of delivery. Among babies who were delivered vaginally, identify those who have a length of stay of over 2 days. Among babies delivered via Cesarean Section, identify those who have a length of stay of over 4 days. (Use Z38.00 to identify vaginal births, and Z38.01 to identify Cesarean births. Z-codes are found in patient discharge data. Neonatal length of stay is defined as the date of discharge minus the date of birth).

c) Moderate complications requiring a prolonged length of stay: Among the infants identified in step b, identify those with moderate complications (Use Patient discharge Diagnosis data, examining both primary and other diagnosis and procedure fields for ICD-10 codes identifying specific moderate complications that require a prolonged length of stay for inclusion in the numerator. See Table 11.46 thru Table 11.53)

d) Prolonged neonatal Length of Stay that Exceeds 5 days: In the remaining population, identify babies who have a prolonged length of stay that exceeds 5 days. (Use Patient Discharge Diagnosis Data to determine Length of Stay. Neonatal length of stay is defined as the date of discharge minus the date of birth).

e) Exclude infants with jaundice or social indications: Among babies identified as having a length of stay that exceeds 5 days, exclude those who have jaundice or are in hospital for social indications such as adoption or foster care. (See Table 11.33 thru Table 11.35 in the excel spread sheet in S.2b for jaundice and social exclusion codes)

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

The denominator is comprised of singleton, live born babies who are at least 37.0 weeks of gestation, and over 2500g in birth weight. The denominator excludes most serious fetal conditions that are "preexisting" (present before labor), including prematurity, multiple gestations, poor fetal growth, congenital malformations, genetic disorders, other specified fetal and maternal conditions and infants exposed to maternal drug use in-utero. The final denominator population consists of babies who are expected to do well following labor and delivery and go home routinely with their mothers.

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).

Step 1: Identify and include singleton, inborn, live births (Use Patient discharge Diagnosis data, specifically diagnosis Codes Z38.00 or Z38.01).

Step 2: Identify and include babies with birth weight >= 2500g. (Use ICD10 codes for low birth weight, birth certificate or EMR).

Step 3: Identify and include full term babies, >=37 weeks gestation (Use ICD10 codes or birth certificate variable called best obstetric estimate of gestational age or EMR data).

Step 4: In less than 1% of cases, the best obstetric estimate of gestation age is missing. In these cases, use LMP-based gestational age to identify full term infants. (Use birth certificate or Patient Discharge data).

Step 5: If both sources of gestational age are missing, include only infants who are over 3000g, as they are more likely to be full term.

**Note: List of ICD-10 codes with individual descriptors is available in the Measure Specifications in S2b above and on our web-page as an excel file

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

a) Babies not born in hospitals are excluded as this is a hospital quality performance measure

b) Babies who are part of multiple gestation pregnancies are excluded.

c) Premature infants (babies born before 37 weeks gestational age) are excluded

d) Low birth weight babies (<=2500g) are excluded

e) Babies with congenital malformations and genetic diseases are excluded

f) Babies with pre-existing fetal conditions such as IUGR are excluded

g) Babies who were exposed to maternal drug use in-utero are excluded

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.)*

a)Babies not born in hospitals are excluded as this is a hospital quality performance measure (Exclude all other live birth codes other than Z38.00 and Z38.01)

b)Babies who are part of multiple gestation pregnancies are excluded.

c)Premature infants (babies born before 37 weeks gestational age) are excluded (use best obstetric estimate of gestational age found in the birth certificate to exclude all infants born before 37 weeks. If best obstetric of gestational age is missing, use the LMP gestational age variable instead to identify infants under 37 weeks)

d)Low birth weight babies (<=2500g) are excluded (Use birth certificate birth weight variable to identify infants under 2500g)

e)Babies with congenital malformations and genetic diseases are excluded (Use ICD-10 codes listed in Table 11.30 to exclude infants with these conditions)

f)Babies with pre-existing fetal conditions such as IUGR are excluded (Use ICD-10 codes listed in Table 31 to exclude infants with these conditions)

g)Babies who were exposed to maternal drug use in-utero are excluded (Use ICD-10 codes listed in Table 32 to exclude infants with these conditions)

The excel document is found in S2b above and on our website.

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.)*

STEP 1: Calculate Denominator Inclusions

a)Identify and include singleton, inborn, live births (Use Patient discharge Diagnosis data, specifically diagnosis Codes V30.00 or V30.01 listed in Appendix 1).

b)Next, identify and include babies with birth weight >= 2500g. (Use birth certificate or Patient Discharge data).

c)Next, identify and include full term babies, >=37 weeks gestation (Use birth certificate variable called best obstetric estimate of gestational age). In less than 1% of cases, the best obstetric estimate of gestation age is missing. In these cases, use LMP-based gestational age to identify full term infants. (Use birth certificate or Patient Discharge data).

d)If both sources of gestational age are missing, include only infants who are over 3000g, as they are more likely to be full term. (Use the birth certificate variable for birth weight).

STEP 2: Calculate Denominator Exclusions

a)In the singleton, full term, population of neonates obtained in Step 1, identify and exclude babies with all congenital malformations and genetic disorders (Use codes listed in Appendix 2, Group A to exclude infants)

b)After congenital malformations and genetic disorders are excluded, further exclude babies with fetal conditions such as IUGR (Use codes listed in Appendix 2, Group B to exclude infants)

c)After babies with congenital malformations, genetic disorders and fetal conditions are excluded, further exclude infants who were exposed to maternal drug use in-utero. (Use codes listed in Appendix 2, Group C to exclude infants).

d)This is the measure's final denominator population

Step 3: Numerator Inclusions: PART A: SEVERE COMPLICATIONS

a)Identify and include Neonatal Deaths (Using patient discharge diagnosis data, specifically the disposition code for death)

b)Identify and include neonatal transfers (Using patient discharge diagnosis data, specifically the disposition code for transfer to a higher level of care)

c)Identify and include babies with "Apgar at 5 minutes" OR "Apgar at 10 minutes" scores of less than 4 (Use Birth certificate or medical record to obtain Apgar scores)

d)Identify and include babies with Severe Morbidities (Use patient discharge diagnosis data, examining both primary and other diagnosis and procedure fields for specific ICD-9 Codes defining an array of specific severe complications. Please refer to Appendix 3, Groups 3A through 3I as the codes are too numerous to include here)

e)Identify and include babies with a Sepsis code and a length of stay that exceeds 4 days (Use patient discharge diagnosis data, examining both primary and other diagnosis fields for the specific ICD-9 code defining sepsis but also requiring a neonatal length of stay of over 4 days. Note that neonatal stay is defined as the date of discharge minus the date of birth).

The neonates identified in Step 3 comprise the "Severe Complications" component of the numerator.

Step 4: Numerator Inclusions: PART B: MODERATE COMPLICATIONS

In the remaining infants (those without severe morbidities), identify and include the following

a)Identify babies with moderate complications that do not require a specific length of stay for inclusion (Use Patient discharge Diagnosis data, examining both primary and other diagnosis and procedure fields for specific ICD-9 codes identifying specific moderate complications (see Appendix 4, Groups A though C)

b)Identify babies with a specified prolonged length of stay stratified by method of delivery. In the population of babies who were delivered vaginally, identify those who have a length of stay of over 2 days. Among babies delivered via Cesarean Section, identify those who have a length of stay of over 4 days.

c)Among babies identified as having a prolonged length of stay (stratified by method of delivery), identify and include those who have moderate complications (Use Patient discharge Diagnosis data, examining both primary and other diagnosis and procedure fields for specific ICD-9 codes identifying specific moderate complications. See Appendix 4, Groups D through H)

d)In the remaining population, identify babies who have a prolonged length of stay that exceeds 5 days. Use Patient Discharge Diagnosis Data to determine Length of Stay

e)Among babies identified as having a length of stay that exceeds 5 days, exclude those who have jaundice or are in hospital for social indications such as adoption or foster care (See Appendix 5 for jaundice and social exclusion codes)

Step 5: Calculation of Unexpected Complications in Term Newborns measure:

Unexpected Newborn Complications (Total): Rate per 100 live births.

(Severe Complications + Moderate Complications/ Final Denominator) x100

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

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

Not applicable. The measure does not require sampling or a survey. This is a major advantage as by using discharge diagnosis files, every hospital can have a large sample (~85% of all births) giving a the most robust assessment of infant outcomes. However, it is recommended that hospitals have at least 200 qualifying cases in the denominator population of this metric (i.e) Full term infants with no pre-existing conditions, malformations, etc described in S.9

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.

Claims

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.

This measure utilizes claims data (hospital discharge diagnosis files) alone or a linked dataset obtained from two separate data sources, patient discharge data and clinical data (gestational age and birthweight) from the EMR, paper records or birth certificate files.

All three approaches have been extensively used in CA, WA and OR.

Patient Discharge Data:

Obtained from the Office of Statewide Planning and Discharge (OSHPD). This dataset does not include data on births from military/naval hospitals as they do not submit data to OSHPD.

Linked to:

Birth Certificate Files:

Obtained from the Center for Health Statistics

The linkage has been used to validate the data available on the claims file and is not needed for use of the measure.This measure utilizes a linked dataset obtained from two separate data sources, patient discharge data and birth certificate files.

Patient Discharge Data:

Obtained from the Office of Statewide Planning and Discharge (OSHPD). This dataset does not include data on births from military/naval hospitals as they do not submit data to OSHPD.

Linked to:

Birth Certificate Files:

Obtained from the Center for Health Statistics

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, Integrated Delivery System, Population : Regional and State

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. This is not a composite measure.

2. Validity – See attached Measure Testing Submission Form

0716_UNC_NQF_testing_attachment_12.13.19_FINAL.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*)**:** NQF 0716
Measure Title: Unexpected Complications in Term Newborns
Date of Submission: Click here to enter a date

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*).

We used a linked dataset of **Patient Discharge Data** obtained from the Office of Statewide Planning and Discharge (OSHPD), State of California. OSHPD datasets do not include data on births from military hospitals. Patient discharge data was linked to **Birth Certificate Files** obtained from California Department of Public Health, Center for Health Information and Statistics.

1.3. What are the dates of the data used in testing? January 1, 2017 to December 31, 2017

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

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.20***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☒ other: regional, integrated delivery system, state	☒ other: regional, integrated delivery system, state

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

All 238 hospitals with maternity services in California in 2017 were included in the testing. Their characteristics are listed below.

Hospital Characteristics	**N (%)**
AAP Neonatal level of care	
Level I	72 (30.3)
Level II	57 (23.9)
Level III	92 (38.7)
Level IV	17 (7.1)
Geographic region	
Central-South Coast	92 (38.7)
Central-North Coast and Northeastern	96 (40.3)
Central Valley, Southern Inland	50 (21.0)
Rural or Urban-Suburban	
Urban-Suburban	205 (86.1)
Rural	33 (13.9)
Average annual delivery volume (livebirths)	
<1,000	72 (30.3)
1,000-2,499	117 (49.2)
>=3,000	49 (20.6)
Hospital ownership	
University, City, County	39 (16.4)

Integrated Health System	29 (12.2)
Private non-profit	128 (53.8)
Private investor	42 (17.6)

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

Testing was performed on 373,763 singleton term newborns without preexisting conditions in 2017. Their characteristics were listed below.

Newborn Characteristics	N (%)
Gestational Age	
37-38	93,132 (24.9)
39-40	247,965 (66.3)
≥ 41	32,561 (8.7)
Unknown (identified by appropriate birth weight of > 3000g)	105 (< 0.1)
Sex	
Male	190,818 (51.1)
Females	182,945 (49.0)
Race	
Non-Hispanic White	103,606 (28.5)
Non-Hispanic Black	18,685 (5.2)
Asian	59,334 (16.3)
Hispanic	178,162 (49.1)
Other	3,348 (0.9)
Method of Delivery	
Cesarean Section	107,550 (28.8)
Vaginal birth	266,213 (71.2)

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.

<u>Reliability Sample.</u> Reliability testing excluded 13 hospitals that contributed fewer than 200 cases to the denominator as based on exclusions recommended in the specifications of this measure. The final testing sample consisted of 225 hospitals and 372,139 singleton term newborns without preexisting conditions.

<u>Validity Samples.</u> We performed four empirical validity tests. The testing samples used for each test are described below.

Testing a & b. We used the same testing sample as described in sections 1.5 and 1.6 (2017 CMQCC linked data with 238 hospitals and 373,763 singleton term newborns without preexisting conditions).

Testing c. We extracted 2014-2017 CMQCC linked data. The number of hospitals and newborns in each time period was specified below.

	2014	2015 Jan to Sep	2016	2017
N of hospital	237	238	238	238
N of all newborns	475,165	344,922	499,919	438,043
N of singleton term newborns without preexisting conditions	402,241	292,642	383,419	373,763

Testing d. Data from 2017 CMQCC active track data were used. CMQCC active track data consisted of selected chart review data elements and newborn patient discharge records that were submitted by CMQCC member hospitals and were linked to birth certificate data in the California Maternal Data Center under CMQCC.

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.

Social risk factors that are available in the CMQCC linked dataset include maternal age, race/ethnicity, education level, gender, gestational age of baby and insurance type. Additionally, our organization obtained data on the American Academy of Pediatrics (AAP) neonatal intensive care level for all research performed by our groups. We also obtained hospital ownership type from California's Office of Statewide Health Planning and Development (OSHPD).

2a2. RELIABILITY TESTING

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

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

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

Reliability testing was performed at the hospital level for 225 hospitals across California who had over 200 cases meeting our denominator inclusion criteria. For purposes of reliability testing, the 2017 California Statewide linked data was analyzed as described in the RAND Corporations "The Reliability of Provider Profiling: A Tutorial" by John L. Adams (RAND Corporation, TR-653-NCQA, 2009). This methodology is specifically recommended by NQF to analyze the reliability of performance for performance measure scores.

Reliability in this context represents the ability of the proposed measure to confidently and accurately distinguish the performance of one entity (hospital) from another. As outlined in the RAND tutorial, "Conceptually, it is the ratio of signal to noise. The signal in this case is the proportion of variability in measured performance that can be explained by real differences in performance." There are 3 main drivers of reliability; sample size, differences between entities (hospitals), and measurement error.

Reliability is estimated using a beta-binomial model, which is appropriate for measuring the reliability of the UNC metric by hospital. The strategy involves fitting a beta-binomial model for the performance metric results. Two parameters (alpha and beta) that define the beta-binomial distribution are generated from the model. From these parameters, the "between hospital variance" was produced. Next, the within hospital variance was generated based on the proportion of affirmative answers. Analyzing the between hospital variance and the within hospital variance generates the reliability for each hospital site.

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*)

The mean reliability score for the 225 hospitals was 0.90. Results are summarized by Hospital deciles in Table 1 below with the means, standard deviations, minimum and maximum reliability statistics and number of hospitals in each decile presented. Individual hospital performance score and reliability statistics are presented in Table 2.

Table 1. Distribution of Unexpected Newborn Complication Reliability Score, summarized by Hospital Deciles.

Rank for Reliability	N of hospitals	Mean	Std Dev	Minimum	Maximum
1	22	0.67	0.06	0.53	0.75
2	23	0.80	0.03	0.75	0.84
3	22	0.87	0.01	0.85	0.88
4	23	0.90	0.01	0.88	0.91
5	22	0.92	0.01	0.91	0.93
6	23	0.94	0.00	0.93	0.95
7	23	0.96	0.00	0.95	0.96
8	22	0.97	0.00	0.96	0.97
9	23	0.97	0.00	0.97	0.98
10	22	0.98	0.00	0.98	0.99

Table 2. Unexpected Newborn Complication Reliability Score, by hospital

Hospital	N of cases in denominator	N of cases in numerator	Performance score (%)	Reliability score
1	5726	86	1.50	0.99
2	3838	39	1.02	0.99
3	3408	47	1.38	0.99
4	5327	116	2.18	0.99
5	3871	64	1.65	0.98
6	3451	53	1.54	0.98
7	2564	30	1.17	0.98
8	468	1	0.21	0.98
9	4336	88	2.03	0.98
10	2920	40	1.37	0.98
11	3315	54	1.63	0.98
12	3032	46	1.52	0.98
13	6910	258	3.73	0.98
14	4235	97	2.29	0.98
15	2235	28	1.25	0.98
16	2793	44	1.58	0.98
17	2568	39	1.52	0.98
18	3396	71	2.09	0.98

19	1508	14	0.93	0.98
20	3625	83	2.29	0.98
21	2390	36	1.51	0.98
22	3799	93	2.45	0.98
23	2200	32	1.45	0.98
24	2181	32	1.47	0.98
25	2336	38	1.63	0.97
26	2700	52	1.93	0.97
27	2145	33	1.54	0.97
28	374	1	0.27	0.97
29	2164	34	1.57	0.97
30	4503	154	3.42	0.97
31	4934	188	3.81	0.97
32	1703	22	1.29	0.97
33	1768	24	1.36	0.97
34	5105	208	4.07	0.97
35	1605	20	1.25	0.97
36	2262	40	1.77	0.97
37	1374	15	1.09	0.97
38	3423	95	2.78	0.97
39	5204	224	4.30	0.97
40	1922	30	1.56	0.97
41	3713	114	3.07	0.97
42	1553	20	1.29	0.97
43	2154	39	1.81	0.97
44	2477	53	2.14	0.97
45	1389	17	1.22	0.97
46	3083	86	2.79	0.97
47	336	1	0.30	0.97
48	2120	41	1.93	0.97
49	3627	122	3.36	0.97
50	2358	51	2.16	0.97
51	1656	25	1.51	0.97
52	2009	37	1.84	0.97
53	3415	110	3.22	0.97
54	2442	57	2.33	0.97
55	724	5	0.69	0.97
56	1603	25	1.56	0.96
57	2570	65	2.53	0.96
58	2038	41	2.01	0.96
59	2485	63	2.54	0.96
60	2873	85	2.96	0.96
61	1254	16	1.28	0.96

62	3943	163	4.13	0.96
63	2245	52	2.32	0.96
64	1745	32	1.83	0.96
65	1198	15	1.25	0.96
66	1643	29	1.77	0.96
67	2056	46	2.24	0.96
68	2319	59	2.54	0.96
69	2399	64	2.67	0.96
70	1555	27	1.74	0.96
71	1268	18	1.42	0.96
72	1723	34	1.97	0.96
73	1746	35	2.00	0.96
74	1180	16	1.36	0.96
75	1973	46	2.33	0.96
76	2663	85	3.19	0.96
77	2743	91	3.32	0.96
78	2989	111	3.71	0.96
79	2679	89	3.32	0.96
80	1722	37	2.15	0.96
81	2387	72	3.02	0.96
82	2083	57	2.74	0.95
83	1704	38	2.23	0.95
84	2126	60	2.82	0.95
85	3089	130	4.21	0.95
86	861	10	1.16	0.95
87	1398	27	1.93	0.95
88	2666	108	4.05	0.95
89	2894	130	4.49	0.95
90	1606	40	2.49	0.95
91	572	5	0.87	0.95
92	1534	37	2.41	0.94
93	1471	34	2.31	0.94
94	1127	20	1.77	0.94
95	1966	63	3.20	0.94
96	2457	101	4.11	0.94
97	1519	38	2.50	0.94
98	1451	35	2.41	0.94
99	1780	53	2.98	0.94
100	1344	30	2.23	0.94
101	2306	90	3.90	0.94
102	2162	79	3.65	0.94
103	1872	61	3.26	0.94
104	1013	18	1.78	0.94

105	2796	142	5.08	0.94
106	1339	32	2.39	0.94
107	2659	131	4.93	0.94
108	1755	57	3.25	0.94
109	1203	27	2.24	0.94
110	2014	77	3.82	0.94
111	1435	39	2.72	0.93
112	826	13	1.57	0.93
113	1567	48	3.06	0.93
114	2111	89	4.22	0.93
115	932	17	1.82	0.93
116	2278	108	4.74	0.93
117	887	16	1.80	0.93
118	1303	35	2.69	0.93
119	1939	79	4.07	0.93
120	821	14	1.71	0.93
121	2788	169	6.06	0.93
122	2265	111	4.90	0.93
123	2030	89	4.38	0.93
124	651	9	1.38	0.93
125	2680	163	6.08	0.93
126	1411	44	3.12	0.92
127	848	16	1.89	0.92
128	1209	34	2.81	0.92
129	1154	31	2.69	0.92
130	1950	92	4.72	0.92
131	1064	27	2.54	0.92
132	1167	34	2.91	0.92
133	1712	75	4.38	0.91
134	348	3	0.86	0.91
135	1244	40	3.22	0.91
136	897	21	2.34	0.91
137	802	17	2.12	0.91
138	1514	62	4.10	0.91
139	1451	57	3.93	0.91
140	2812	224	7.97	0.91
141	2204	136	6.17	0.91
142	861	20	2.32	0.91
143	508	7	1.38	0.91
144	1063	32	3.01	0.91
145	721	15	2.08	0.90
146	879	23	2.62	0.90
147	1281	51	3.98	0.90

148	513	8	1.56	0.90
149	2685	240	8.94	0.90
150	695	15	2.16	0.90
151	916	27	2.95	0.89
152	1910	124	6.49	0.89
153	998	33	3.31	0.89
154	1428	70	4.90	0.89
155	456	7	1.54	0.89
156	886	27	3.05	0.89
157	1302	62	4.76	0.88
158	577	12	2.08	0.88
159	522	10	1.92	0.88
160	1043	41	3.93	0.88
161	1917	144	7.51	0.88
162	1106	47	4.25	0.88
163	863	29	3.36	0.87
164	1221	59	4.83	0.87
165	1134	51	4.50	0.87
166	641	16	2.50	0.87
167	532	11	2.07	0.87
168	1161	54	4.65	0.87
169	935	36	3.85	0.87
170	969	39	4.02	0.87
171	658	18	2.74	0.87
172	1142	57	4.99	0.86
173	647	18	2.78	0.86
174	793	28	3.53	0.86
175	989	45	4.55	0.86
176	570	15	2.63	0.85
177	702	24	3.42	0.85
178	428	9	2.10	0.85
179	1193	74	6.20	0.84
180	598	18	3.01	0.84
181	652	22	3.37	0.84
182	1811	184	10.16	0.84
183	676	24	3.55	0.84
184	643	22	3.42	0.84
185	454	11	2.42	0.83
186	1604	148	9.23	0.83
187	272	4	1.47	0.83
188	1099	71	6.46	0.83
189	578	19	3.29	0.83
190	1442	126	8.74	0.83

191	799	38	4.76	0.82
192	291	5	1.72	0.82
193	595	24	4.03	0.80
194	439	13	2.96	0.80
195	436	13	2.98	0.80
196	257	5	1.95	0.78
197	316	8	2.53	0.77
198	327	9	2.75	0.76
199	878	70	7.97	0.76
200	472	20	4.24	0.75
201	261	6	2.30	0.75
202	544	27	4.96	0.75
203	481	21	4.37	0.75
204	236	5	2.12	0.75
205	412	17	4.13	0.73
206	374	15	4.01	0.72
207	963	108	11.21	0.72
208	305	10	3.28	0.72
209	395	17	4.30	0.72
210	371	15	4.04	0.72
211	482	26	5.39	0.71
212	535	35	6.54	0.70
213	603	46	7.63	0.69
214	317	13	4.10	0.68
215	287	11	3.83	0.67
216	613	56	9.14	0.66
217	440	30	6.82	0.65
218	358	21	5.87	0.63
219	249	10	4.02	0.63
220	266	12	4.51	0.62
221	314	17	5.41	0.62
222	265	14	5.28	0.58
223	262	14	5.34	0.58
224	220	10	4.55	0.57
225	362	34	9.39	0.53

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?*)

Reliability scores can vary from 0.0 to 1.0, where a score of zero implies that all variation is attributable to measurement error (noise) whereas a reliability of 1.0 implies that the variation is caused by real differences in performance across hospitals. According to the RAND report, reliability scores of 0.7-0.8 are considered

acceptable for drawing conclusions across hospitals, and a score of 0.9 are considered sufficient to see differences between individuals.

The mean reliability score of our metric in the testing sample was 0.90, which is very good. Mean decile hospital reliability scores were above 0.7 in 9/10 deciles, and 7/10 deciles have mean scores of over 0.90. The reliability testing results show that variation in scores is caused by real differences in performance across the hospitals and is not due to measurement error.

2b1. VALIDITY TESTING

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

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

☒ **Performance measure score**

☒ **Empirical validity testing**

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

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

I: EMPIRICAL VALIDITY TESTING OF MEASURE SCORE.

We examined several options for construct validity testing. NICU admission as a comparator was not our first choice because of data indicating large variation (40 -fold!) among hospitals for NICU admission for term infants (Schulman J etal, 2018) that appeared to be primarily related to bed availability. Hospital cost (adjusted from charges) has been used as a reliable marker of morbidity including recent studies with Severer Maternal Morbidity (ChenHY etal, 2018). Similarly, length of stay can be used an independent marker of degree of illness.(Snowden CP etal, 2013). We tested Unexpected Newborn Complications measure against each of these three alternative measures for construct validity.

Schulman J, Braun D, Lee HC, Profit J, Duenas G, Bennett MV, Dimand RJ, Jocson M, Gould JB.

Association Between Neonatal Intensive Care Unit Admission Rates and Illness Acuity.

JAMA Pediatr. 2018 Jan 1;172(1):17-23.

Chen HY, Chauhan SP, Blackwell SC. Severe Maternal Morbidity and Hospital Cost among Hospitalized Deliveries in the United States. Am J Perinatol. 2018 Nov;35(13):1287-1296.

Snowden CP, Prentis J, Jacques B, Anderson H, Manas D, Jones D, Trenell M. Cardiorespiratory fitness predicts mortality and hospital length of stay after major elective surgery in older people. Ann Surg. 2013 Jun;257(6):999-1004.

Testing a. We conducted a patient-level analysis to evaluate the association between Unexpected Newborn Complications measure and newborn length of stay and newborn hospital cost. The study population included 373,763 singleton term newborns without preexisting conditions in 238 California hospitals in 2017. Newborns that were deceased (N=17) or transferred to another facility (N=3,306) were excluded, which resulted in 370,440 newborns in this analysis. We performed univariate analysis of the mean, standard deviation, median, and interquartile range (IQR). We used Wilcoxon two-sample test to test whether the distribution of newborn

length of stay and newborn hospital cost was different between those Unexpected Complications and those without.

Testing b. By using the same testing sample as described in testing a, we conducted a hospital-level analysis to assess Pearson Correlation Coefficient between hospital rate of unexpected newborn complications and hospital average newborn length of stay and hospital average newborn cost. we measured newborn hospital cost by summing charges for the newborn from the hospital discharge record and then converting the charges to cost using a refined cost-to-charge ratio approach, based on California hospitals, as described in X Xu, et al., 2017.

Xu X, Lee HC, Lin H, Lundsberg LS, Pettker CM, Lipkind HS, Illuzzi JL. Hospital variation in cost of childbirth and contributing factors: a cross-sectional study. BJOG. 2018 Jun;125(7):829-839.

Testing c. We compared the rate of unexpected newborn complications in the ICD-9 period (2014 and 2015 Jan to Sep) and in the ICD-10 period (2016 and 2017).

Testing d. CMQCC member hospitals had actively submitted their chart review data on NICU admission to the Maternal Data Center within CMQCC. In 2017, 49 hospitals had over 95% monthly review rate on their newborn records, which resulted in 80,852 records. We assessed Pearson Correlation Coefficient between hospital rate of unexpected newborn complications and hospital rate of NICU admission among these records.

II: SYSTEMATIC ASSESSMENT OF FACE VALIDITY:

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

I: EMPIRICAL VALIDITY TESTING OF MEASURE SCORE.

Testing a. The mean and median length of stay and newborn hospital cost were markedly higher among term babies with unexpected complications compared to those without complications. Results from Wilcoxin two-sample test showed that there is a statistically significant difference between the underlying distributions of length of stay and newborn hospital cost between the two groups ($P < 0.01$, Table 3 & Table 4). The exceptionally large average differences between babies with and without Unexpected Newborn Complications for both length of stay and cost is strong evidence for the validity of this measure.

Table 34. Distribution of newborn length of stay, by Unexpected Newborn Complications (yes/no)

Unexpected Newborn Complications	N	Mean	Std	Median	Q1	Q3	*P*-value from Wilcoxin two-sample test
Yes	8,278	5.2	3.4	4	3	7	< 0.01
No	362,162	1.9	0.8	2	1	2	

Table 4. Distribution of newborn hospital cost, by Unexpected Newborn Complications (yes/no)*

Unexpected Newborn Complications	N	Mean	Std	Median	Q1	Q3	*P*-value from Wilcoxin two-sample test
Yes	6,894	$10,751.5	$13,683.5	$7,942.5	$4,115.6	$12,890.8	< 0.01
No	295,558	$1,197.2	$1,306.6	$920.3	$639.4	$1,321.5	

*30 of the 238 California hospitals were not included from the cost analysis because of missing charges data.

Testing b. Hospital average newborn length of stay and hospital average newborn cost were both positively associated with hospital rate of unexpected newborn complications, with a correlation coefficient of around 0.4 ($P < 0.01$) (Table 5).

Table 5. Pearson Correlation coefficient (r) for the relation of unexpected newborn complication rate, average newborn length of stay, and average hospital costs in 238 California hospitals, 2017

	Unexpected Newborn Complications Rate	
	r	***P*-value**
Average newborn length of stay	0.41	< 0.01
Average newborn hospital cost	0.37	< 0.01

Testing c. Around 85% of newborns were eligible to be included in the denominator of the measure of unexpected newborn complications, and the inclusion rate was stable from the ICD-9 to the ICD-10 period (Table 6). The rate of total Unexpected Newborn Complications also remained unchanged during the same period. In response to experience from the first years of use, in 2016 we also adjusted LOS requirements for several of the codes for severe UNC which resulted in a mild reduction in the rates of severe UNC and a corresponding rise in moderate UNC.

Table 6. Rate of unexpected newborn complications in 238 California hospitals, 2014-2017

ICD Version	**ICD-9**		**ICD-9**		**ICD-10**		**ICD-10**	
Year	**2014**		**2015 Jan to Sep**		**2016**		**2017**	
N of newborns	**475,165**		**344,922**		**449,919**		**438,043**	
N of newborns in measure denominator	**402,241**		**292,642**		**383,419**		**373,763**	
% of newborns in measure denominator	**84.7**		**84.8**		**85.2**		**85.3**	
Unexpected Newborn Complications	**N**	**Rate (%)**	**N**	**Rate (%)**	**N**	**Rate (%)**	**N**	**Rate (%)**
Severe	8562	2.1	5885	2.0	6702	1.7	6628	1.8
Moderate	4703	1.2	3335	1.1	5094	1.3	4973	1.3
Total	13265	3.3	9220	3.2	11796	3.1	11601	3.1

*Rate per 1,000

Testing d. The correlation coefficient of the rate of unexpected newborn complication and NICU admission was 0.64, with a *P*-value of < 0.01. This illustrated that simple NICU admissions among term babies does correlate with Unexpected Newborn Complications but has room for improvement as suggested by the Schulman article referenced above. Also of note is that NICU admission is not routinely available in administrative data sets unless they contain revenue codes.

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?*)

I: EMPIRICAL VALIDITY TESTING OF MEASURE SCORE.

The measure of unexpected newborn complications was developed to capture term babies with unexpected health outcomes, which indicated the need of prolonged and more intensive health care. Therefore, we expected that this measure would be correlated with NICU admission, newborn length of stay, and newborn hospital cost. Results from testing a, b, and d proved our hypothesis, and the positive but not the strongest correlations indicate the measure is not providing redundant information. Indeed, there is evidence that NICU admissions includes a sizable number of term infants that are not critically ill and this rate varies from facility to facility. This would suggest that UNC is superior to simple NICU admission as a measure of quality in this population.

In addition, results from testing c showed that the inclusion rate for this measure and the rate of the measure itself remained unchanged from the ICD-9 to the ICD-10 period, indicating that the changes of the codes (ICD9 to ICD10 and measure maintenance tweaks) had minimal effect on the measure.

2b2. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

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*)

To identify singleton, hospital-born, babies born in California in 2017, we used a linked dataset of patient discharge data linked to vital statistics birth certificate records. Naval and military hospitals were not included, as they do not submit patient discharge data to the Office of Statewide Health Planning and Development. Our analysis was limited to identifying morbidity occurring during the birth admission only. We did not track re-admissions.

After limiting our dataset to the above admissions, the following exclusions were analyzed for frequency and variability across hospitals included in our analysis:

These criteria are also used for the Joint Commission PC-06 measure: Unexpected Complications in Term Newborns Version 2020A2.

- ICD-10-CM Principal Diagnosis Code or ICD-10-CM Other Diagnosis Codes for birth weight < 2500g as defined in Appendix A, Table 11.12, 11.13, 11.14, 11.15, 11.16, 11.20 OR Birth Weight < 2500g.
- Patients who are not term or with < 37 weeks gestation completed.
- Patients whose term status or gestational age is missing and birthweight < 3000 gm.
- ICD-10-CM Principal Diagnosis Code or ICD-10-CM Other Diagnosis Codes for congenital malformations and genetic diseases as defined in Appendix A, Table 11.30 Congenital Malformations.
- ICD-10-CM Principal Diagnosis Code or ICD-10-CM Other Diagnosis Codes for pre-existing fetal conditions as defined in Appendix A, Table 11.31 Fetal Conditions.
- ICD-10-CM Principal Diagnosis Code or ICD-10-CM Other Diagnosis Codes for maternal drug use exposure in-utero as defined in Appendix A, Table 11.32 Maternal Drug Use.

Note: Gestational age was defined based on the “Best Obstetric Estimate of Gestational Age” from the birth certificates. In the few cases that had missing values for Best Obstetric Estimate of Gestational Age, we used Gestational Age according to the Last Menstrual Period instead.

Exclusions were performed in a hierarchical manner in the order listed above. The exclusion steps described above are also detailed on the Joint Commission website (https://manual.jointcommission.org/releases/TJC2020A2/MIF0393.html).

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*)

We examined the overall frequencies and proportions of the admissions excluded for each exclusion criterion in the 2017 California patient discharge-birth certificate linked data. After limiting the dataset to singleton, hospital-born babies, our initial cohort included 431,029 births. The final cohorts, after applying the additional exclusions described below included 373,763 births.

Categories are not mutually exclusive and statistical analyses present below were performed on the initial cohorts of singleton, hospital-born babies.

Table 7. Frequency of denominator exclusions for the measure of unexpected newborn complications (431,029 singleton, hospital-born babies in 238 hospitals in California, 2017)

		Distribution across hospitals (in percentiles)		
Exclusions	**N (%)**	**25th**	**50th**	**75th**
Low Birth Weight Exclusion (<2500g)	22,786 (5.3)	3.6	4.7	6.0
Low Gestational age exclusion (<37 weeks) using Obstetrician's best estimate of Gestational Age	29,977 (7.0)	4.8	6.3	7.6
Missing gestational age and birth weight < 3000g	66 (< 0.1)	< 0.1	< 0.1	< 0.1
Congenital Malformation Exclusion	16,808 (3.9)	2.1	3.1	4.3
Non-malformation Exclusion	34,479 (8.0)	5.1	6.8	8.8
Maternal Drug Use Exclusion	4,285 (1.0)	0.3	0.7	1.8

2b2.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

The overall frequency of the exclusions is low (n=64,280, 14.7%). The distribution of exclusions across hospitals is as expected. However, we feel that each of the exclusions is absolutely necessary and should be retained as their inclusion would bias the performance results by including a subset of preterm, low-birth weight babies many of whom are already in poor health before birth. The inclusion of these children would bias the results and confound true performance differences in obstetric and neonatal quality at the hospital level. The metric's intended denominator population is healthy, full term neonates who are "expected to go home routinely" but unexpectedly experience adverse events.

Gestational week Exclusion:

Rationale: The measure excludes premature babies (those born before 37 weeks of gestation).

Birth weight Exclusion:

Rationale: The measure excludes low birth weight infants as they may be premature, small for gestational age, or experienced intra-uterine growth restriction. Many low birth babies were also premature and would have been excluded anyway.

Congenital Malformation Exclusion:

Rationale: Babies with congenital malformations are excluded as they are generally not healthy and may have a myriad of conditions that require additional medical treatment soon after birth and later in life.

Pre-Existing Conditions Exclusion:

Rationale: Babies who are light for dates (small gestational age), experienced fetal growth retardation, who were affected by placenta previa, as well as fetuses affected by hemolytic disease due to Rh isoimmunization or hydrops were also excluded as they are not considered healthy and require additional medical treatment.

Maternal Drug Use Exclusion:

Rationale: Babies whose mothers used drugs during pregnancy were excluded as these infants would have suffered withdrawal, had longer neonatal lengths of stay and could have other health problems associated with exposure to drugs in-utero.

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 2b5.

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 Single low risk strata 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.

RATIONALE

In the context of healthcare performance assessment, the purpose of a risk adjustment model is to reduce bias due to case mix characteristics present at the start of care (in this case the admission for birth of the baby). This measure is not risk-adjusted but rather risk-stratified, using a series of exclusions (described above) to identify a standard low-risk population. When constructing the measure, the exclusion criteria were chosen to ensure that the target population would be healthy, term babies with no pre-existing complications, thus reducing bias due to case mix complications. Babies more at risk for experiencing adverse outcomes (premature babies, low birth weight infants, babies with congenital malformations, exposure to maternal substance use and other pre-existing conditions) were excluded from the target population. The rationale for each of the exclusions is outlined in Question 2b2.3.

The National Quality Forum prefers that measures are not risk adjusted for patient factors that could possibly obscure disparities (namely age, sex and socioeconomic status). Therefore, we did not adjust for sex or insurance status of the newborns. We chose not to adjust for gestational age (within the term, 37-43 weeks of gestation, population) recognizing that some morbidities are more prevalent at different gestational ages because timing of labor induction is part of obstetric practice. In short, we did not want to mask morbidities

resulting from early elective delivery practices (under 39 weeks of gestational age) or non-interventional practices in some hospitals (who do not induce women who are over 41 weeks pregnant, thus increasing the risk of stillbirth and morbidity in post term infants).

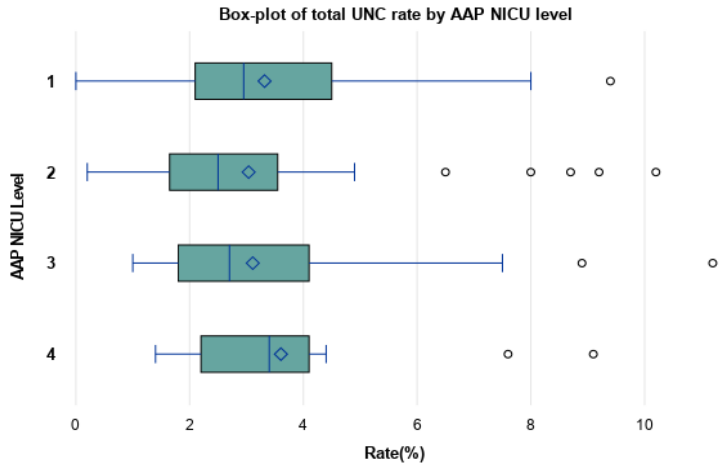
Variables related to quality of care are purposely not included in risk models for performance measures used to assess quality. Risk adjustment should not mask or adjust for the very factors that are driving the differences in neonatal health outcomes at hospitals across California. Accordingly, we did not adjust for a hospital's neonatal intensive care unit level, birth volume, ownership status, teaching status or number of maternal-fetal care specialists. The list of exclusions account for most conditions that have been linked to social risk factors such as preterm birth and poor fetal growth (small-for-dates infants) so we did not further assess social risk.

ANALYSES

To investigate whether unaccounted case mix variation could affect the measure we also examined distribution of hospital rates of UNC by APP NICU Levels. As shown in Table 8 and the related box-plot, there is extensive comparability and overlap of results among all four levels confirming the absence for need for further risk adjustment.

Table 8. Distribution of hospital rate of unexpected newborn complications in 234 California hospitals, 2017, by AAP NICU level

AAP NICU Level	N of hospitals	Mean	Std	Min	Q1	Median	Q3	Max
I	70	3.3	1.8	0.0	2.1	3.0	4.5	9.4
II	56	3.0	2.1	0.2	1.6	2.5	3.5	10.2
III	91	3.1	1.8	1.0	1.8	2.7	4.1	11.2
IV	17	3.6	2.1	1.4	2.2	3.4	4.1	9.1



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)**

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*)

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

Not Applicable

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:

Single low-risk strata; see Table 8 above in **2b3.2**

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

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

Not Applicable

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

2b4.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the*

steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)

To examine differences in performance, we calculated 95% confidence intervals for the unadjusted metric results for all eligible hospitals. If a hospital's confidence interval did not include the California state mean (mean of the unexpected newborn complication results for all eligible hospitals in California) then the hospital was identified as statistically significantly better or worse than the California state average.

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*)

We excluded 4 hospitals with less than 100 denominators in the measure of unexpected newborn complications, which resulted in 234 hospitals with 373,534 singleton term newborns without preexisting conditions. The distribution of hospital rate was shown in Table 9.

Table 9. Distribution of hospital rate of unexpected newborn complications in 234 California hospitals, 2017

Mean	SD	Minimum	10th percentile	25th percentile	Median	75th percentile	90th percentile	Maximum
3.2	1.9	0.0	1.4	1.8	2.8	4.1	5.4	11.2

Using the approach described in 2b4.1, 46 hospitals (19.7%) of 234 California hospitals were rated as statistically significantly higher (worse) than the state mean (i.e. the lower limit of hospital's 95% confidence interval was > 3.2) and 89 hospitals (38.0%) were identified as statistically significantly lower (better) than the state mean (i.e. the upper limit of hospital's 95% confidence interval was < 3.2). Another 99 hospitals were either higher or lower than the state mean but their results were not statistically significant.

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?*)

This measure is able to detect hospitals with better and worse than average performance. Hospitals that were identified in 2b4.2. above as statistically significantly better or worse than the state average had scores that were at least 20% lower or 15% higher than the state mean, which we consider a meaningful difference in performance.

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

2b6.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)
A few variables are essential to the calculation of this metric. They are outlined below, and we have outlined safeguards to deal with the rare instances in which they are missing.

Missing Z38.00 or Z38.01 codes: These codes present in patient discharge data identify in-hospital singleton births. If these variables are missing, we are unable to identify in-hospital births and would have to exclude the case from the measure.

Missing Gestational Age (Using Best Obstetric Estimate of Gestational Age): In less than 1% of cases in linked administrative data, the best obstetric estimate of gestation age (from birth certificate data) is missing. In these cases, we use the LMP-based gestational age to identify full term infants. If both sources of gestational age are missing, we only include infants who are over 3000g, as they are more likely to be full term.

Missing Birth Weight: Birth weight is missing in less than 0.05% of cases in administrative data. If birth weight is missing, we have 2 sources of gestational age to be able to include an eligible baby into the metric.

Under-coding of diagnoses: The hierarchical construction of the numerator of this metric provides several double-checks to minimize the chance of missing a newborn with an unexpected newborn complication. If a truly sick newborn is missed at one stage of the metric, it can be captured in the next levels of the metric. We also require a length of stay for certain conditions to protect against over coding and under coding certain complications. Finally, if no complications are coded at all, a baby with a length of stay of over 5 days that does not have social reasons for remaining in hospital (adoption/foster care) or jaundice will be included in the metric. Of course, this measure will not identify newborn complications in infants born at home or at birthing centers where discharge files with ICD-10 codes are not submitted.

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*)
This was answered in the question above (2b6.1) for the missing birth weight, gestational age and in-hospital birth variables. In the cases where diagnoses are under or over coded, **hospitals perform routine audits of hospital charts** if they find that their unexpected newborn complication rate is being driven by a particular diagnosis. Furthermore, we have worked with individual hospitals who have been able to identify coding practices to change resulting in improvement of their measure scores. One tertiary hospital changed its use of

CPAP in keeping with regional norms, after it was found that CPAP was being mistakenly being over-coded in newborn records.

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

Missing data elements are too miniscule to have any effect on the measure. As stated in our responses to questions 2b6.1 and 2b6.2, the unexpected newborn complications measure has built-in checks to account for possible missing data, as well as under and over coding. It is highly unlikely that our measure will miss including an eligible infant.

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.

Coding/abstraction performed by someone other than person obtaining original information (E.g., DRG, ICD-9 codes on claims, 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**.

ALL data elements are in defined fields in a combination of 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).

We are exploring with The Joint Commission the feasibility of this measure being developed as an eCQM. We do not see any strong barriers. At this point it is more a matter of bandwidth and resources. While we have not completed a formal survey for eCQM Feasibility, we believe that the key data elements (ICD10 diagnosis codes; ICD10 procedure codes; date of birth; date of discharge) all meet the criteria for availability, accuracy, standardized data, and routinely captured in usual workflow.

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.

Coding Practices 1: Over and under coding: We learned that coding practices do vary for some ICD-10 codes with some hospitals being "over exuberant" in their coding and others clearly under-coding existing complications. The new specifications attempt to balance this issue by requiring that many codes for Moderate Complications additionally have an infant length of stay that exceeds the typical maternal postpartum length of stay (>2 days for a vaginal birth and >4 days for a cesarean birth). This requirement significantly reduces the number of infants identified but validates that these babies had significant morbidity. Conversely, some babies had very long neonatal length of stay without any codes to account for it, suggesting the possibility of under-coding. Our expert panel identified two categories of prolonged neonatal length of stay that were not medically serious and could be excluded from this consideration, namely neonatal jaundice typically treated with Bili-Lights, and social disruption for homelessness or foster care. We found that a number of babies with septicemia had short length of stay indicating that it was not likely severe, therefore we added a requirement for a length of stay of at least 5 days to be included among Severe Complications.

Coding Practices 2: We have noted that in some higher-level hospitals coding practices varied from standard to achieve billing targets. The most common example was noted in tertiary facilities where the code for CPAP was routinely used for bag and mask resuscitation in the Delivery Room even when used for less than one minute. In response, the codes for CPAP were moved to the category that required a postpartum LOS that was longer than the typical LOS (See above).

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)*.

No fees or licensing required.

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 The Joint Commission https://www.jointcommission.org/-/media/tjc/documents/measurement/march-2020-performance-measurement-reporting-optional-and-extended-timelines.pdf Public Health/Disease Surveillance California Maternal Quality Care https://www.cmqcc.org/focus-areas/quality-metrics/unexpected-complications-term-newborns Regulatory and Accreditation Programs The Joint Commission https://www.jointcommission.org/measurement/measures/perinatal-care/ Professional Certification or Recognition Program BC BS Blue Distinction for Maternity Care https://www.bcbs.com/sites/default/files/file-attachments/page/Evaluation_Components_2020_Maternity_Care.pdf Quality Improvement (external benchmarking to organizations) California Maternal Quality Care Collaboartive National Perinatal Information Center https://www.cmqcc.org/focus-areas/quality-metrics/unexpected-complications-term-newborns http://www.npic.org/index.php Quality Improvement (Internal to the specific organization) California Maternal Quality Care https://www.cmqcc.org/focus-areas/quality-metrics/unexpected-complications-term-newborns National Perinatal Information Center http://www.npic.org/index.php Safe Deliveries Roadmap Collaborative: Washington State Hospital Association (WSHA) http://www.wsha.org/quality-safety/projects/safe-deliveries/ Oregon Perinatal Collaborative: Oregon Health Care Quality Corporation (QCorp) http://q-corp.org/maternity-care

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

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

Current user 1: California Maternal Quality Care Collaborative
Purpose: Quality improvement with bench marking, internal quality improvement
Scope: All 235 California birthing hospitals representing just over 460,000 births annually.
Current User 2: National Perinatal Information Center(NPIC)
Purpose: Quality improvement with bench marking, internal quality improvement
Scope: All 85 NPIC member hospitals across the US representing 360,000 births annually
Current User 3:
Safe Deliveries Roadmap Collaborative: Washington State Hospital Association (WSHA)

Purpose: quality improvement with bench marking, internal quality improvement
Scope: 35 hospitals in Washington, Alaska and Montana representing approximately 175,000 newborns annually
Current User 4:
Oregon Perinatal Collaborative: Oregon Health Care Quality Corporation (QCorp)
Purpose: Quality improvement with bench marking, internal quality improvement
Scope: 14 hospitals in Oregon representing approximately 55,000 newborns annually.
Current User 5
The Joint Commission: PC-06 Unexpected Newborn Complications
Purpose: Accreditation with bench marking, internal quality improvement, public reporting
Scope: All TJC participating hospitals with maternity services over 200 birth/year (>2500 US hosptials)
Current User 6:
Blue Distinction-Blue Cross Blue Shield Association: Blue Distinction® Centers for Maternity Care Program
Purpose: Certification and Recognition
Scope: All Blue Cross-Blue Shield participating hospitals (>2500 US hospitals)

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

This measure is being used by The Joint Commission for accreditation and will shortly be ready for public reporting. Our philosophy has been to provide hospitals with their own data for 2 to 3 years and then begin public reporting of hospital results. This measure has been shared state-wide on the California Maternal Data Center. Participating hospitals in California, Washington,Oregon, Alaska and Montana as well as 80+ National Perinatal Information Center (NPIC) member hospitals across the nation have access to this metric and have specific benchmarks and trends to follow their internal data. In addition, we provide extensive drill-down analysis for hospitals to understand why their rate may be elevated.
The Unexpected Newborn Complications metric specifications are available free of cost to any interested user on our CMQCC website.

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.*)

As noted above this measure is currently being used for a variety of accountability applications. In addition, as of Jan 2020, it is being collected by the Joint Commission with the intent of public release later this year. CMQCC provides hospital-level metrics for public reporting to the California Hospital Accountability and Reporting Task force (CHART), supported by the California Health Care Foundation.

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.

Over 300 hospitals with birthing facilities (in California, Oregon, Washington and Alaska (over 650,000 births) are members of the California Maternal Data Center (MDC). The MDC provides real-time benchmarking and analytic tools including the ability to drill down into their denominator cases. These tools allow facilities to better understand the drivers for their rates and focus their improvement efforts. CMQCC supports multiple user groups to garner feedback on measures and the data center analytic tools. In this process, we have received informative data about coding practices that maybe contrary to coding guidelines. Where possible we have provided coding education with our collaborating coding specialists but in some cases we have actually adjusted the specifications to more closely follow actual coding practice. The National Perinatal Center with its large collection of hospitals (85 facilities and 360,00 annual births) has also provided on-going feedback. This

has been especially useful as we adjust the ICD10 codes every fall with the additions and modifications provided by CMS.

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.

CMQCC provides monthly updates to all member hospitals that includes multiple benchmarking against like size, like NICU level, county and perinatal region. Further analyses include breakdown to severe and moderate UNC as well as categories of complications (e.g. respiratory, infection, birth injury, neurologic). We have provided a series of webinars on the measure which have been widely attended. The CMQCC MDC has extensive user support including clinical support for understanding and interpretation of the measure. The CMQCC website also contains further information. NPIC has also run seminars and provided written summary information for its members. The Joint Commission has also provided written information and staged several webinars jointly with CMQCC.

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 CMQCC Maternal Data Center (MDC) has multiple user support paths including clinical support for understanding and interpretation of the measure. Feedback is obtained from over 300 facilities of all sizes using written help tickets, user groups and forums. NOIC has also generated comments and questions form its user groups. More recently, The Joint Commission Perinatal has generated questions and comments as it has been rolled out for national application.

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

Through our user groups and webinars and other sources of feedback we have created an long set of FAQ´s with extensive discussion that we have published on our website and has been widely shard by the Joint Commission.(https://www.cmqcc.org/sites/default/files/Unexpected_Newborn_Complications_FAQs_Current.pdf)

Questions have included with discussions: Is there a target rate? How to use the measure? Do hospitals caring for higher risk patients have higher rates of UNC? Is a high UNC rate due to coding or care? Will ta case that has been transferred be counted for the delivery or the receiving hospital? Are extramural deliveries excluded from UNC? How are transfers of newborns with known anomalies counted? What about a transfer of a newborn at the request of an insurer?

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

NPIC and TJC and out partner organizations in WA and OR have provided feedback from their member hospitals and are similar to that from CA. The types of questions and concerns are reflected in the FAQs summarized above and on the link:
https://www.cmqcc.org/sites/default/files/Unexpected_Newborn_Complications_FAQs_Current.pdf

Both organizations have been very positive about the usefulness of the measure and are promoting its wide application. Blue Cross Blue Shield Association is now including UNC in its Blue Distinction program for perinatal center recognition.

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.

The feedback has been instrumental in adjusting the measure over its 5+ years of widespread usage. We have been able to tweak the codes used and adjust their place in categories requiring a minimum LOS. While this occurred mostly int he first year and in the first year after transition to ICD10 we are still getting useful feedback on specific codes. A broader topic that we have received feedback on is the categorization of "transfer to higher level of care" as an UNC criteria. This ends up being an important issue as the it is a driver

of UNC for small hospitals and does represent both a separation of baby from family as well as significant complications that are typically poorly coded by the lower level facility. The measure was not modified for this feedback but further feedback and QI efforts were initiated instead (see 4b2.1 below).

Improvement

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

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

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

This measure was used as an important balancing measure for a large scale (56 hospitals, 119,000 annual births) QI collaborative to reduce primary cesarean birth. UNC did not worsen when Cesarean rates were reduced by 24% but actually improved by 8%. (Main EK, Chang S, Cape V, Sakowski C, Smith H, Vasher J. Safety Assessment of a Large-Scale Improvement Collaborative to Reduce Nulliparous Cesarean Delivery Rates. Obstet Gynecol 2019;133: Apr;133(4):613-623). This was followed by state-wide efforts to improve maternity care and lower Cesarean rates and the severe UNC rates also improved (from 18.2 per thousand in 2015 to 16.1 in 2018 and 14.3 in preliminary Q1-2 2019 data. These data represent 235 hospitals and ~460,00 annual births.

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.

No unintended consequences have been noted. We are especially promoting that this metric be used as a balancing measure to identify unintended consequences of other measures such as Cesarean Section reduction. It has been very useful in this regard to allay worries about reduction.

One unexpected finding is the very wide variation among Level 1 nurseries of transferring out term babies without any prior conditions. This varies from 0.5% to 6% among California hospitals and in a similar study done in Northern New England (VT, NH and ME). In both settings, follow-up interviews identified that the practice was not due to patient characteristics but rather to staff level of experience and confidence caring for babies with mild respiratory complications (primarily the unease for even a few hours of observation for transient tachypnea). This turn separated the baby from the family without clear need. Both quality collaboratives have identified this as an education and training opportunity.

One additional issue that has appeared largely in California hospitals that have sub-contracted their NICUs to a CHildren´s Hospital. As the Children´s Hospital has a separate licensure, all admissions to the their NICU, even for minor reasons, are treated technically as a transfer meeting an UNC criteria. As it is very important for the obstetric hospital to understand the outcomes of its infants we have worked with the joint operations to allow for the OB facility to report all of the ICD10 codes of babies they have "transferred "to the in-facility NICU.

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

We were not expecting to find that as the California NTSV Cesarean rate declined in a large state-wide collaborative that the Severe UNC rate would also improve. (Main EK, Chang S, Cape V, Sakowski C, Smith H,

Vasher J. Safety Assessment of a Large-Scale Improvement Collaborative to Reduce Nulliparous Cesarean Delivery Rates. Obstet Gynecol 2019;133: Apr;133(4):613-623)

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.

There is no other currently endorsed measure in this topic area. A formerly endorsed NQF measure (NQF # 0474 Birth Trauma -Injury to the Neonate) would have been considered a "competing measure" as it conceptually addressed the same measure focus and target population. It suffered from over coding issues with several ICD codes dominating the measure that were ambiguous (e.g. "Other birth injuries NOS"). This remains an issue for ICD-10. For that and other reasons, that measure was "un-endorsed". Furthermore that measure was focused only on physical birth injuries while our measure identifies a much broader range of neonatal morbidities that are a consequence of labor and delivery.

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.)

NQF 0474 Birth trauma Rate (AHRQ PSI#17) was submitted in 2012 and not endorsed. It provides a very limited window into term morbidities from the single perspective of birth trauma. There are many other morbidities in term infants that are much more common and important to quantify as several of them are severe and can have long lasting implications well into childhood and beyond. We feel our measure is superior to NQF 0474 for the following reasons:

• We examine a much broader range of adverse events including deaths, transfers, low Apgar scores and a wide range of severe and moderate conditions. Including hypoxic encephalopathy, very low Apgar scores, and respiratory distress in term infants.
• After consulting neonatologists, pediatricians and obstetricians about the severity of certain conditions and how to quantify and group conditions appropriately, we are confident that our measure differentiates between severe and moderate morbidity.
• Our measure factors in neonatal length of stay, which is an important indicator in assessing whether an infant is truly severely ill or not. For example, an infant may have a diagnosis code for neonatal sepsis (a very serious newborn complication) but if the neonatal LOS was only 2 days (and no death or transfer) se We also include method of delivery and its impact on length of stay, as infants delivered via Cesarean section generally stay in hospital for four days and infants born vaginally stay for two days or less. We exclude conditions like jaundice and social factors that cause infants to have longer neonatal lengths of stay.
• Our exclusions ensure that our denominator (target) population truly does consist of healthy term newborns by excluding preterm infants, low birth weight babies, congenital malformations, babies subjected to maternal drug use and other preexisting conditions.
• Our measure allows hospitals to drill down into sub-measures of morbidity such as respiratory complications, neurological complications and infections to determine what is driving their unexpected newborn complication rate.
•The larger incidence of conditions in our measure compared to the NQF birth injury measure allows for much better statistical analysis and discrimination. Furthermore, our measure is currently used to evaluate over 1 million births in multiple states and hospitals across the US (corresponding to approximately 25% of all US births).

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.

Attachment **Attachment:** Appendix_face_validity-637219911930854945.docx

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): California Maternal Quality Care Collaborative

Co.2 Point of Contact: Elliott, Main, main@cmqcc.org, 415-992-2252-

Co.3 Measure Developer if different from Measure Steward: California Maternal Quality Care Collaborative

Co.4 Point of Contact: Elliott, Main, main@cmqcc.org, 415-992-2252-

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.

CMQCC members: Elliott Main, MD; Anisha Abreo MPH, Debra Bingham RN DrPH; Kathryn Melsop, MS

CPQCC members: Terri Slagle, MD and Richard Powers, MD (both Neonatologists long active in QI research)

MQI members: Kimberly Gregory, M.D., MPH; Lisa Korst, MD PhD; Moshe Freedman, PhD; Sonal Shah, MPH; Michael Lu, MD MPH.

The entire team reviewed and discussed the concepts and ICD9 codes. MQI did the first pass of the data analysis, CMQCC did subsequent. Testing with focus groups and with other organizations was done by CMQCC.

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2009

Ad.3 Month and Year of most recent revision: 07, 2013

Ad.4 What is your frequency for review/update of this measure? Every 2-3 years

Ad.5 When is the next scheduled review/update for this measure? 05, 2020

Ad.6 Copyright statement: This measure will be in the public domain.

Ad.7 Disclaimers:

Ad.8 Additional Information/Comments: