
CBE ID

0716

Title

Unexpected Newborn Complications in Term Infants

Project

Primary Prevention

Endorsement Status

Endorsed

Is Under Review

No

Next Maintenance Cycle

Spring 2030

Previous Endorsement Cycle

Spring 2025

Initial Endorsement

Sun, 01/16/2011 - 15:37

Steward

California Maternal Quality Care Collaborative

1.0 New or Maintenance

Maintenance

1.1 Measure Structure

Single Measure

1.3 Electronic Clinical Quality Measure (eCQM)

No

1.6 Measure Description

This measure identifies moderate and severe complications among term infants who enter labor without preexisting fetal conditions (morbidity outcome measure). The scope of the measure includes both maternal and neonatal care and is appropriate for all levels of hospital care and is collected for a full year time period. The measure is designed to highlight processes of both obstetric and neonatal care that can be improved while recognizing that not all complications can be avoided. As an infant outcome measure it can also serve as a balancing measure for interventions targeting maternal outcomes such as reducing cesarean birth. It has been in use by the California Maternal Quality Care Collaborative in 3 states (California, Oregon, and Washington) for over 10 years and over 5 years by The Joint Commission (PC-06).

1.7 Measure Type

Outcome

1.8 Level of Analysis

Facility

1.9 Care Setting

Hospital: Inpatient

1.10 Measure 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 is a complete lack of metrics that assess the health outcomes and guide quality improvement activities for 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 maternal measures such as Nulliparous Term Singleton Vertex (NTSV) Cesarean, third- and fourth-degree lacerations, early elective delivery rates, and neonatal practices such as admissions to rule out sepsis for all infants exposed to chorioamnionitis. The purpose of a balancing measure is to guard against any unanticipated or unintended consequences of quality improvement activities for these measures.

1.11 Measure Webpage

<https://manual.jointcommission.org/releases/TJC2025B/MIF0393.html>

1.13 Data Dictionary

Attached

1.13a Attach Data Dictionary

[CBE 0716 data dictionary APR2025_Battelle.xlsx](#)

1.14 Numerator

Unexpected newborn complications including neonatal death, transfer to a higher level of care, birth trauma, hypoxia/asphyxia, shock and resuscitation, respiratory complications, neurologic complications, infection, and long length of stay

1.14a Numerator Details

In the full-term neonatal population that excluded low birth weight infants, those with congenital malformations and pre-existing conditions, and those exposed to maternal drug use, infants 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 data, specifically the disposition code for death)
- b) Neonatal transfers (Use patient discharge data, specifically the disposition code for transfer to a higher level of care)
- c) Severe morbidities: (Use patient discharge data, examining both primary and other diagnosis and procedure fields for ICD-10-CM codes defining an array of specific severe complications. Please refer to Tables 11.36 through 11.45 with the specific ICD-10-CM codes and descriptors listed in excel document in 1.13a.
- d) Sepsis with a neonatal length of stay that exceeds 4 days (Use patient discharge data, examining both primary and other diagnosis fields for the specific ICD-10-CM codes defining sepsis. Note that neonatal length of 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 data, examining both primary and other diagnosis and procedure fields for ICD-10-CM codes identifying specific moderate complications (see Table 11.46 through Table 11.48 for the specific ICD-10-CM codes and descriptors listed in excel document in 1.13a).
- 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 ICD-10-CM code Z38.00 to identify vaginal births, and Z38.01 to identify cesarean births. Z-codes are found in patient discharge data, both primary and other diagnosis fields. 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 data, examining both primary and other diagnosis and procedure fields for ICD-10-CM codes identifying specific moderate complications that require a prolonged length of stay for inclusion in the numerator. See Tables 11.49 through Table 11.53 for the specific ICD-10-CM codes and descriptors listed in excel document in 1.13a).
- d) Prolonged neonatal length of stay that exceeds 5 days: In the remaining population, identify

newborns who have a prolonged length of stay that exceeds 5 days. (Use patient discharge 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 the hospital for social indications such as adoption or foster care. (Use patient discharge data, examining both primary and other diagnosis and procedure fields for ICD-10-CM codes for jaundice and social exclusion codes. See Tables 11.33 through Table 11.35 in the Excel spreadsheet in 1.13a).

1.15 Denominator

Term, singleton live births with birth weight $\geq 2,500$ grams, and without pre-existing conditions including congenital malformations, genetic disorders, other fetal-placental conditions, intrauterine growth restriction, and those exposed to maternal drug use or withdrawal symptoms

1.15a Denominator Details

The denominator is comprised of singleton, live-born infants who are at least 37.0 weeks of gestation and 2,500g or over 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.

Step 1: Identify and include singleton, inborn, live births (Use patient discharge data, examining both primary and other diagnosis fields for ICD-10-CM code Z38.00 or Z38.01).

Step 2: Identify and include babies with birth weight $\geq 2,500$ g. (Use EMR or birth certificate data for low birth weight).

Step 3: Identify and include full-term babies, defined as ≥ 37.0 weeks of gestation (Use EMR or birth certificate data for obstetric estimate of gestational age).

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

Step 5: If both sources of gestational age are missing, include only infants who are over 3,000g, as they are more likely to be full term.

1.15b Denominator Exclusions

a) Infants not born in hospitals

- b) Infants who are part of multiple gestation pregnancies.
- c) Premature infants (born before 37 weeks gestational age)
- d) Low birth weight Infants (< 2,500g)
- e) Infants with congenital malformations and genetic diseases
- f) Infants with pre-existing fetal conditions such as IUGR
- g) Infants who were exposed to maternal drug use in-utero

1.15c Denominator Exclusions Details

- a) Infants who are not born in the hospital or are part of multiple gestation pregnancies (exclude infants with no ICD-10-CM codes for single liveborn newborn Z38.00 and Z38.01)
- b) Low birth weight infants (< 2,500g) are excluded (Use EMR or birth certificate data)
- c) Premature infants (infants born before 37.0 weeks' gestational age) are excluded (use the best obstetric estimate of gestational age found in EMR or birth certificate data to exclude all infants born before 37 weeks. If the best obstetric gestational age is missing, use the LMP gestational age instead.
- d) Babies with congenital malformations and genetic diseases are excluded (Use ICD-10-CM codes listed in Table 11.30 to exclude infants with these conditions)
- e) Babies with pre-existing fetal conditions such as IUGR are excluded (Use ICD-10-CM codes listed in Table 31 to exclude infants with these conditions)
- f) Babies who were exposed to maternal drug use in utero are excluded (Use ICD-10-CM codes listed in Table 32 to exclude infants with these conditions)

****Note:** List of ICD-10-CM codes with individual descriptors is available in the Data Dictionary in 1.13a.

1.15d Age Group

Other

1.15e Age Range in Years

Newborns (0 years)

1.16 Type of Score

Rate/proportion

1.17 Measure Score Interpretation

Other

1.17a Other Measure Score Interpretation

Better quality = Lower score but placed into 3 categories described in Risk Adjustment section 5.4.1. Severe UNC is considered a much more important marker of QI concern while Moderate UNC still has QI opportunities.

1.18 Calculation of Measure Score

STEP 1: Calculate Denominator Inclusions

- a) Identify and include singleton, inborn, live births (Use patient discharge data, examining both primary and other diagnosis fields for ICD-10-CM codes Z38.00 or Z38.01).
- b) Next, identify and include babies with birth weight $\geq 2,500\text{g}$. (Use EMR or birth certificate data).
- c) Next, identify and include full-term babies, ≥ 37 weeks gestation (Use the best obstetric estimate of gestational age from EMR or birth certificate data). In less than 1% of cases, the best obstetric estimate of gestational age is missing. In these cases, use LMP-based gestational age to identify full-term infants. (Use EMR or birth certificate data).
- d) If both sources of gestational age are missing, include only infants who are over 3,000g, as they are more likely to be full term. (Use EMR or birth certificate data for birthweight.)

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 Table 11.30, Congenital Malformations, 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 Table 11.31, Fetal Conditions, to exclude infants)
- c) After infants 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 Table 11.32, Maternal Drug Use, 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 (Use patient discharge data, specifically the disposition code for death)
- b) Identify and include neonatal transfers (Use patient discharge data, specifically the disposition code for transfer to a higher level of care)
- c) Identify and include babies with severe morbidities (Use patient discharge data, examining both primary and other diagnosis and procedure fields for specific ICD-10-CM codes defining an array of specific severe complications. Please refer to Tables 11.36 Severe Birth Trauma, 11.37 Severe Hypoxia/Asphyxia, 11.38 Severe Shock and Resuscitation, 11.39 Neonatal Severe Respiratory Complications, 11.40 Neonatal Severe Infection, 11.41 Neonatal Severe Neurological Complications, 11.42 Severe Shock and Resuscitation Procedures, 11.43 Neonatal Severe Respiratory Procedures, and 11.44 Neonatal Severe Neurological Procedures)
- d) Identify and include babies with a sepsis ICD-10-CM diagnosis code and a length of stay that exceeds 4 days (Use patient discharge data, examining both primary and other diagnosis fields for the specific ICD-10-CM codes defining sepsis (Please refer to Table 11.45, Neonatal Severe Septicemia) but also requiring a neonatal length of stay of over 4 days. Note that neonatal length of stay is defined as the date of discharge minus the date of birth).

The neonates identified in Step 3 Part A comprise the “Severe Complications” component of the numerator.

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 data, examining both primary and other diagnosis and procedure fields for specific ICD-10-CM codes identifying specific moderate complications (Tables 11.46 Moderate Birth Trauma, 11.47 Moderate Respiratory Complications, 11.48 Moderate Respiratory Complications Procedures).
- b) Identify babies with a specified prolonged length of stay stratified by method of delivery. In newborns who were delivered vaginally, identify those who have a length of stay of over 2 days. Among newborns delivered via cesarean section, identify those who have a length of stay of over 4 days.

- c) Among newborns identified as having a prolonged length of stay (stratified by method of delivery), identify and include those who have moderate complications (Use patient discharge data, examining both primary and other diagnosis and procedure fields for specific ICD-10-CM codes identifying specific moderate complications. Tables 11.49 Moderate Birth Trauma with LOS, 11.50 Moderate Respiratory Complications with LOS, 11.51 Moderate Neurological Complications with LOS Procedures, 11.52 Moderate Respiratory Complications with LOS Procedures, and 11.53 Moderate Infection with LOS)
- d) In the remaining population, identify newborns who have a prolonged length of stay that exceeds 5 days. Use patient discharge data to determine length of stay.
- e) Among newborns 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 (Tables 11.33 Neonatal Jaundice, 11.34 Phototherapy, and 11.35 Social Indications)

PART C: MODIFICATIONS OF TRANSFERS TO A HIGHER LEVEL OF CARE (new for 2025)

For infants identified in Step 3 Part A-b as being transferred to a higher level of care, assign them to the numerator as follows:

- a) Moderate Complications component: if the infant has no severe complications defined in Step 3 Part A-c and Part A-d, but does have at least one moderate complication listed in Step 3 Part B-a through Part B-e, re-classify them to the “Moderate Complications” component of the numerator.
- b) All other transferred infants remained in the “Severe Complications” component of the numerator.

(Among the 1,381 infants identified in Step 3 Part A-b as transfers to a higher level of care across 197 hospitals in 2022, only 107 (7.8%) were re-classified into the “Moderate Complications” component of the numerator. The re-classification opportunity aims to motivate hospitals to improve the completeness of coding for infants who are transferred out.)

Step 4: Calculation of Unexpected Complications in Term Newborns measure:

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

$(\text{Severe Complications} + \text{Moderate Complications} / \text{Final Denominator}) \times 100$

1.18a Attach measure score calculation diagram

[PC-06 \(v2025B\) algorithm.pdf](#)

1.19 Measure Stratification Details

This year, we have added a stratification based on hospitals' neonatal level of care defined by the American Academy of Pediatrics (AAP). We separated AAP Level I (Basic; no NICU) facilities from AAP Levels II-IV facilities (NICUs at increasing sophistication). AAP Level I facilities are small, typically in rural areas, and have a statistically different distribution of UNC rates than hospitals with NICUs. This is driven by their higher rates of neonatal transfers to facilities with a NICU. These transfers account for the majority of severe UNC cases for many of the Level I hospitals. But there is a marked variation (>5-fold) among Level I facilities, highlighting a Quality Improvement (QI) opportunity. A neonatal transfer to a NICU, often at a distance, of a term infant without preexisting conditions is a major anxiety for the mother and family. Unfortunately, the Level I hospital patient discharge diagnosis data for these transfers are often very limited, and rarely is the reason for the transfer present in the ICD-10-CM codes. Certainly, some infants without preexisting conditions require transfer in every Level I setting, but the rate should not vary from <0.5 percent to over 5%. As the reasons for their elevated UNC rates differ from infants in AAP Level II-IV facilities (and hence the QI opportunities) and the distribution of hospital rates was so different, we felt there was strong reason to stratify the results. The box and whiskers plots in Supplementary Figures S1-1 and S1-2 illustrate the distribution of severe UNC across the four AAP Level NICUs. Median severe UNC rates ranged from 2.0% in Level I NICUs to 0.9% in Level III and IV, with overlapping interquartile ranges across all NICU levels (Supplementary Table S1-1). A Kruskal-Wallis test showed statistically significant differences in severe UNC rate distributions across NICU levels ($P < 0.01$). However, a post-hoc analysis limited to Levels II-IV hospitals indicated less variation among higher-level NICUs ($P = 0.06$). Median total UNC rates ranged from 2.5% in Level I NICUs to 3.0% in Level IV NICUs, with no statistically significant differences in distribution across the four NICUs ($P = 0.42$) (Supplementary Table S1-2).

1.20 Types of Data Sources

Administrative Data, Electronic Health Records, Other

1.20a Other Data Source

Birth certificate data

1.20d Format: Other Data Source

Digital

1.21a Data Collection Tool URL(s)

<http://example.com>

1.25 Data Source Details

This measure utilizes either patient discharge data alone or a linked dataset that combines patient discharge data with clinical information (gestational age and birthweight) from EMR or birth

certificate records.

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

Patient discharge Data:

Obtained from the California Department of Health Care Access and Information (HCAI). This dataset does not include data on births from military/naval hospitals, as they do not submit data to HCAI.

Linked to:

Birth certificate Data:

Obtained from the Center for Health Statistics

1.26 Minimum Sample Size

The measure does not require sampling or a survey. This is a major advantage as by using patient discharge data, every hospital can have a large sample (~85% of all births), giving 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, as described in 1.15c.

2.1 Attach Logic Model

[2.1 logic model_updated_Final.docx](#)

2.2 Evidence of Measure Importance

Ten recent studies have used Unexpected Newborn Complications as either a key outcome or important balancing measure focused on improving obstetric practice, and many 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 the 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 a 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.

(6) Dombrowski (2020) examined the risk of trial of labor among mothers with two prior cesarean deliveries ($N=42,771$). Among these women, trial of labor was rarely attempted (1,228, 2.9%) and was successful in 39.4% of attempts. Trial of labor in this population was not associated with an increase in maternal morbidity but was associated with a modest increase in severe neonatal complications (OR 1.78, 95% CI 1.04-3.04).

(7) Rosenstein (2021) evaluated a statewide California comprehensive quality improvement collaborative to reduce cesarean births (published in JAMA). The study included nearly 7.6 million NTSV births and showed a highly significant statewide reduction of the cesarean rate to below the

HP20230 target of 23.6%. The statewide rate of severe UNC (used as a balancing measure) did not worsen and significantly improved (2.1% to 1.5%).

(8) Panelli (2021) examined maternal and neonatal morbidity after attempted operative vaginal delivery. Successful operative vaginal delivery was associated with reduced severe maternal morbidity (adjusted odds ratio, 0.55; 95% confidence interval, 0.39-0.78) without a difference in severe unexpected neonatal morbidity (adjusted odds ratio, 0.99; 95% confidence interval, 0.78-1.26). In contrast, failed operative vaginal delivery was associated with increased severe maternal morbidity (adjusted odds ratio, 2.14; 95% confidence interval, 1.20-3.82) and severe unexpected neonatal morbidity (adjusted odds ratio, 1.78; 95% confidence interval, 1.09-2.86).

(9) Rosenstein (2024) further explored the statewide cesarean reduction program and compared 65 hospitals that reduced their first-birth cesarean rates to meet the HP 2030 target to 72 hospitals that did not meet the national cesarean target. In the hospitals that met the target, the severe UNC rate fell from 3.6 to 2.8% ($p < 0.05$), while in the hospital that did not meet the target, there was no significant reduction (2.8% to 2.5%).

(10) Fineberg (2024) examined in detail the significant reduction of cesarean rates in 3 Sacramento area hospitals during a local effort to reduce cesarean deliveries, focusing on management of the second stage and elective induction of labor. Active phase labor management was more important than labor induction in lowering the cesarean rate. There was no change in the rate of unexpected newborn complications after the interventions.

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 Table 5.3-3 in Section 5.3, Validity Testing of the Measure, illustrate a good correlation between Unexpected Newborn Complications and neonatal hospital charge (a 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 narrower range (cord blood gasses or birth injuries) or an overly broad range (NICU admissions) of concerning neonatal conditions than the more tightly configured Unexpected Newborn Complications. Moving beyond correlations with other neonatal outcomes, several studies described here showed improvements in Unexpected Newborn Complications following large-scale quality improvement/safety initiatives. This illustrates the actionability of Unexpected Newborn Complications. (See Section 6.2.4 for data showing a state-wide improvement).

Reference list:

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.

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.

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.

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.

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.

Dombrowski M, Illuzzi JL, Reddy UM, Lipkind HS, Lee HC, Lin H, Lundsberg LS, Xu X. Trial of Labor After Two Prior Cesarean Deliveries: Patient and Hospital Characteristics and Birth Outcomes. *Obstet Gynecol*. 2020 Jul;136(1):109-117.

Rosenstein MG, Chang SC, Sakowski C, Markow C, Teleki S, Lang L, Logan J, Cape V, Main EK. Hospital Quality Improvement Interventions, Statewide Policy Initiatives, and Rates of Cesarean Delivery for Nulliparous, Term, Singleton, Vertex Births in California. *JAMA*. 2021 Apr 27;325(16):1631-1639.

Panelli DM, Leonard SA, Joudi N, Girsan AI, Judy AE, El-Sayed YY, Gilbert WM, Lyell DJ. Severe maternal and neonatal morbidity after attempted operative vaginal delivery. *Am J Obstet Gynecol* MFM. 2021 May;3(3):100339.

Rosenstein MC, Chang S-C, Tucker CM, Sakowski C, Leonard SA, Main EK. Evaluation of Statewide Program to Reduce Cesarean Deliveries Among Nulliparous Individuals With Singleton

Pregnancies at Term Gestation in Vertex Presentation. Obstet Gynecol 2024 144(4):p 507-515.

Fineberg AE, Harley K, Lahiff M, Main EK. The relative impact of labor induction versus improved labor management: Before and after the ARRIVE (a randomized trial of induction vs. expectant management) trial. Birth. 2024 Dec;51(4):719-727.

2.4 Performance Gap

We used 2022 California birth certificate-linked patient discharge data to assess performance gaps in the unexpected newborn complications (UNC) measure. Initially, we identified 331,546 low-risk births from 222 hospitals in California. For the analysis presented in this report, we excluded 13 hospitals that exhibited artificially elevated UNC rates due to newborns transferred to NICUs within the same facility but under different licenses, an issue we have addressed in recent years. Additionally, we excluded hospitals with fewer than 200 UNC-eligible births in 2022. After these exclusions, the final analysis included 301,114 term newborns delivered at 197 hospitals.

Table 1. Performance Scores by Decile

	Performance Gap												
	Overall	Minimum	Decile_1	Decile_2	Decile_3	Decile_4	Decile_5	Decile_6	Decile_7	Decile_8	Decile_9	Decile_10	Maximum
Mean Performance Score	2.83	0.48	1.01	1.50	1.92	2.25	2.51	2.70	2.96	3.33	4.20	5.98	9.85
N of Entities	197	1	19	20	20	20	19	20	20	20	20	19	1
N of Persons / Encounters / Episodes	301114	208	19041	37878	38226	33947	26286	36165	29607	35039	25204	19721	274

2.4a Attach Performance Gap Results

[performance_gap_attachment.docx](#)

2.6 Meaningfulness to Target Population

This measure was evaluated for meaningfulness by a group of 22 experts including neonatologists, obstetricians, nursing leaders and childcare advocates. The summary question was to assess the statement “This measure, as specified, provides an accurate reflection of quality and can be used to distinguish the quality of obstetric and neonatal care at the hospital level”. The rating scale had 5 levels (1-5) with the following narrative anchors: 1=Strongly Disagree; 2= Somewhat Disagree; 3=Neutral; 4=Somewhat Agree; 5= Strongly Agree. Out of the 22 participants, 19 (86%) gave the measure a rating of 5, thereby strongly agreeing that the scores from the measure as specified would provide a useful reflection of quality. The mean rating was very high 4.82/5 and none of the participants disagreed with the statement.

There has been strong interest in the use of this measure to help provide a more rounded view of hospital maternity care beyond rates of cesarean deliveries and maternal morbidity to add a

focused term infant morbidity measure. To that end, the measure has been adopted as a national measure by The Joint Commission as PC-06. It has been used by the multiorganization state-wide perinatal quality collaboratives in CA, OR and WA for over 10 years to support perinatal quality improvement with QI projects in both neonatal and maternal care. The patient and family advisory groups in these organizations have been particularly supportive of its inclusion. US News & World Report has included it in their annual assessment of hospitals with maternity services.

3.1 Contributions Towards Closing Care Gaps

Evidence of Known Disparities: Glazer KB et al. (Pediatrics 2021) examined Total UNC in 40 New York City hospitals and noted a differential rate for Black and Hispanic women that persisted after multivariable logistic regression analysis (adjusted odds ratio [aOR]: 1.5, 95% confidence interval [CI]: 1.3-1.9; aOR: 1.2, 95% CI: 1.1-1.4, respectively). However, they also noted that “Black and Hispanic women were more likely to deliver in hospitals with high complication rates than were white or Asian American women. Findings implicate hospital quality in contributing to preventable newborn health disparities among low-risk, term births.

In comparison, the California analysis of patients from 197 hospitals using GEE Poisson regression modeling did not find a differential rate for Black and Hispanic women for Total UNC (adjusted risk ratio [RR]: 0.98, 95% CI: 0.88-1.09; aRR: 0.90, 95% CI: 0.85-0.95, respectively) (Supplementary Table S3-1). As in New York City, hospitals themselves appeared to be the important factor in driving the rate of UNC. This analysis is discussed in detail in the Methodology section below.

Methodology: We accessed stratum-specific severe and total UNC rates by maternal age at delivery, maternal race and ethnicity, payment for delivery, and urbanicity/rurality of the delivering hospitals. We performed GEE Poisson regression models with robust error estimates to calculate relative risk (RR) and 95% confidence interval (CI) for severe and total UNC across these social-contextual variables. The adjusted models included the following variables: maternal age at delivery, maternal pre-pregnancy BMI, maternal race/ethnicity, payment for delivery, prenatal care, nulliparous, infant birth weight > 4000 grams, gestation age, infant sex, induction of labor, mode of delivery, rural/urban location of the hospital and maternal comorbidities including chronic hypertension, preeclampsia, gestational diabetes, and diabetes. This approach accounted for the clustering of outcomes within hospitals and provided robust estimates of disparities, adjusted for relevant clinical risk factors and maternal demographics.

Results/Interpretation: These analyses are shown in Supplementary Table S3-1 and indicate that any modest observed differences in patient-level UNC for race/ethnicity and age are not significant when risk-adjusted for the other factors listed above. However, there is an elevated risk for rural hospitals, which will be discussed in detail in Section 5.4.1, Risk Adjustment on the topic of stratification by hospital AAP Level.

As described in Section 5.4.1, the emphasis for interpretation of the UNC measure has shifted from comparing raw rates to categorizing hospital performance as follows: “Within expected range” ($\leq Q3$); “Alert” (from $Q3$ to $Q3 + 1.5 \text{ IQR}$), and “Outlier” (above $Q3 + 1.5 \text{ IQR}$). Our analysis demonstrated that adjusting for individual risk factors, including age and payment, resulted in minimal changes in hospital rates and no change in performance groups. We also examined hospitals with the highest number of Black patients and compared the proportion that fell into “Alert” and “Outlier” categories to all other hospitals. Among the 25 hospitals with the highest number of Black infants meeting UNC inclusion criteria, 1 fell into the “Outlier” category (4.0%) and 3 fell into the “Alert” category (12.0%), comparable to overall rates of 6.2% (Outliers) and 14.5% (Alert). These findings suggest that there is not a need for risk adjustment for race, and the focus should be on hospital QI for all parturients and infants.

Turning to the effect of Medicaid births on UNC rates, we examined the rates of Total and Severe UNC in hospitals with the highest numbers of Medicaid births. Among the highest 10 facilities, none had Total UNC rate or Severe UNC in the Alert or Outlier categories. In the top 40 facilities for Medicaid births, for Total UNC, 10% were in the Alert and 10% in the outlier categories; for Severe UNC, there were none in the Alert category and 5% in the Outlier category. These rates are similar to the rates seen in the population as a whole: Total UNC-16% in Alert and 8% in Outlier; Severe UNC-14% in Alert and 8% in Outlier. These data suggest that a high number of Medicaid births does not in itself disadvantage a hospital for either Total or Severe UNC.

See Section 5.4.1 for a full discussion of risk adjustment. Importantly, infants more at risk for experiencing adverse outcomes (premature babies, low birth weight infants, infants with congenital malformations, exposure to maternal substance use and other pre-existing conditions) were excluded from the target population. Thus, many factors that could drive the need for risk adjustment were excluded a priori.

Anticipated Impact: In California, we do provide all participating hospitals with the opportunity to examine all measures by race and ethnicity as part of their QI process. Should a disparity arise, the cases can be examined to address general and specific issues. This is in concordance with the Glazer article (New York City) noted above. Their conclusion included the following statement: “Quality improvement targeting routine obstetric and neonatal care is critical for equity in perinatal outcomes.” This measure is the only current measure that targets the care of newborns without preconditions that account for 85% of the neonatal population.

Reference list:

Glazer KB, Zeitlin J, Egorova NN, Janevic T, Balbierz A, Hebert PL, Howell EA. Hospital Quality of Care and Racial and Ethnic Disparities in Unexpected Newborn Complications. *Pediatrics*. 2021 Sep;148(3):e2020024091. doi: 10.1542/peds.2020-024091.

4.1a Data Structure and Availability

All required data elements for this measure are routinely generated during standard clinical care and documentation processes. Specifically, gestational age and birth weight are consistently captured in structured electronic fields within the EHR and reported in administrative data sets such as Vital Records and Patient Discharge Diagnosis (PDD) data (ICD-10 codes). Recent studies have demonstrated very high accuracy for ICD-10 codes for gestational age, particularly in distinguishing term from preterm births. Diagnoses, procedures, and lengths of stay are similarly captured in structured electronic fields within PDD data.

Missing data is rare, occurring in fewer than 0.1% of newborns for both birth weight and gestational age in birth certificate data. In instances where gestational age is missing, newborns with a birth weight of 3,000 grams or more are included as the denominator inclusions.

A source of potential inaccuracies arises primarily from variations in coding practices across hospitals, including potential over- or under-coding of complications. We learned in the first years of deployment that coding practices do vary for some ICD codes, with some hospitals being “over exuberant” in their coding and others clearly under-coding existing complications. The specifications attempt to balance this issue by requiring that many codes for 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. We found that a number of babies with septicemia had a short length of stay (2-3 days), indicating that it was not truly sepsis (and therefore not a severe UNC). Therefore, we added a requirement for a length of stay of at least 5 days to be included among Severe Complications. 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.

Since the last submission, no changes have occurred to the measure specifications that impact data structure or availability.

4.1b Implementation Costs and Burden

The implementation of this measure incurs minimal costs since it leverages existing datasets almost exclusively Newborn Discharge File (EMR and Patient Discharge Data, Vital Records only are tapped as a back-up for missing data where available), which hospitals and health systems routinely submit to state and national reporting systems. There is no additional data entry or significant modifications to clinician workflow required, as all utilized data elements are byproducts of standard clinical documentation.

Previous barriers encountered included initial inconsistencies in coding practices, which have been addressed by refined measure specifications incorporating length-of-stay criteria (discussed in section 4.1a). An additional barrier arises in a small number of hospitals with NICUs within the same facility as obstetrics but licensed by a different entity. These hospitals face challenges in ensuring accurate documentation and tracking of ICD-10 diagnosis and procedure codes for newborns transferred internally, as they are reported as being transferred to a different facility. There is a subset of maternity facilities in California (N=13) for which this is a practice, and with effort, work arounds have been developed to link mother and infant outcomes together. This has been reported to occur in other states, but at a much lower frequency, as the financial incentives are different for such licensure arrangements. It is important for these hospitals to adopt mechanisms to track newborn outcomes within their internal NICU to ensure accurate measure calculations and the ability to perform perinatal quality improvement activities.

4.1c Confidentiality

Patient confidentiality is maintained through secure handling of existing datasets (Birth Certificate and Patient Discharge Data), which are routinely managed following strict privacy standards established by state and federal regulations (e.g., HIPAA compliance). Data utilized in this measure are de-identified prior to submission and immediately aggregated for reporting, thus minimizing the risk of confidentiality breaches.

Even when small patient counts occur, standard reporting practices (e.g., suppression or aggregation of small cells) effectively prevent identification of individual patients, further ensuring patient confidentiality.

4.3 Feasibility Informed Final Measure

The feasibility assessment has significantly informed the final measure specifications, guiding necessary adjustments to enhance practicality and accuracy. Initially, efforts to implement the measure using Patient Discharge Diagnosis (PDD) data alone identified substantial under-utilization of gestational age and birth weight ICD codes, which impacted the accurate identification of the target population. In response, the measure specifications were revised to incorporate linkage with Birth Certificate and EMR data, greatly improving the completeness and reliability of critical data elements. This has largely been replaced with the ICD-10 upgrade to gestational age coding, which has proven much more accurate than ICD-9.

Further, variability in hospital coding practices—particularly over- or under-coding of complications—prompted the incorporation of length-of-stay criteria into the measure specifications. This adjustment ensures that moderate and severe complication codes reflect clinically significant conditions, thus reducing potential inaccuracies from inconsistent coding

practices.

Additionally, the feasibility assessment uncovered specific coding challenges among hospitals operating NICUs managed by different entities within the same facility. These hospitals initially had artificially high UNC rates due to higher internal NICU transfer rates and inconsistent capture of diagnoses and procedure codes. Although explicit measure instructions were not introduced, proactive collaboration and targeted outreach helped these hospitals improve their coding and data-capture practices. As a result, their UNC rates have stabilized, accurately reflecting true outcomes.

We and others have noted that the UNC results for small AAP Level I perinatal units are largely driven by transfer of term neonates to higher levels of care. These infants rarely have ICD-10 codes that would explain the need for transfer, or sometimes do not even have any diagnosis at all besides normal birth, making the transfer itself a critical finding. We have also noted substantial variation in the rates (up to 4-fold) of these transfers of term infants without any preexisting conditions. To focus on this issue, we are recommending separate reporting for hospitals without a NICU (AAP Level I) as they appear to have different quality issues than larger facilities.

Collectively, these feasibility-driven modifications ensure the measure can be reliably implemented using routinely available electronic data sources, maintain data accuracy, and minimize hospital burden, thereby enhancing the validity and utility of performance reporting and quality improvement activities.

4.4 Proprietary Information

Not a proprietary measure and no proprietary components

5.1.1 Data Used for Testing

We used a linked dataset of Patient Discharge Data from January 1 to December 31, 2022, obtained from the California Department of Health Care Access and Information (HCAI). HCAI datasets do not include data on births from military hospitals. Patient discharge data was linked to Birth Certificate Files obtained from the California Department of Public Health, Center for Health Information and Statistics.

5.1.1a Dates of Testing Data

Field not required Spring 2025

5.1.2 Differences in Data

None

5.1.3 Characteristics of Measured Entities

We identified 331,546 low-risk births from 222 California hospitals in 2022. We excluded 13 hospitals that exhibited artificially elevated UNC rates due to newborns transferred to NICUs within the same facility but under different licenses, an issue we have addressed in recent years. Additionally, we excluded hospitals with fewer than 200 UNC-eligible births. After these exclusions, the final analysis included 301,114 term newborns delivered at 197 hospitals.

Among the 197 hospitals, 21.8% provided Level I neonatal care, 26.4% Level II, 44.2% Level III, and 7.6% Level IV. Geographically, 40.1% were located in the Central-North Coast and Northeastern regions, 37.6% in the Central-South Coast, and 22.3% in the Central Valley and Southern Inland regions. Most hospitals (90.4%) were situated in urban or suburban areas. In 2022, 45.2% of hospitals delivered between 1,000 and 2,499 live births, 29.9% delivered fewer than 1,000 births, and 24.9% delivered 2,500 or more births. Hospital ownership was primarily private non-profit (49.2%), followed by private investor-owned (18.8%), integrated health systems (14.7%), city/county/district hospitals (12.2%), and university hospitals (5.1%) (Supplementary Table S5.1-1).

5.1.4 Characteristics of Units of the Eligible Population

Approximately 26.2% of low-risk newborns were born to mothers aged 35 years or older. Nearly half (49.1%) were born to mothers of Hispanic ethnicity, followed by 26.9% to non-Hispanic White mothers, 14.5% to Asian mothers, and 4.8% to Black mothers. Gestational age at birth was distributed as follows: 65.1% between 39+0 and 40+6 weeks, 27.5% between 37+0 and 38+6 weeks, and 7.4% at or beyond 41+0 weeks. Cesarean deliveries accounted for 28.1% of births. The infant male-to-female ratio was 1.04 (Supplementary Table S5.1-2).

5.2.1 Level(s) of Reliability Testing Conducted

Accountable entity level (i.e., measure score) (e.g., signal-to-noise analysis)

5.2.2 Method(s) of Reliability Testing

Reliability testing was conducted following the methods described in the RAND Corporation's publication "The Reliability of Provider Profiling: A Tutorial" by John L. Adams (RAND Corporation, TR-653-NCQA, 2009).

Reliability evaluates a measure's ability to consistently and accurately differentiate performance among hospitals. It is defined as the ratio of signal (actual performance variability between hospitals) to noise (random measurement variability).

The three main components that inform reliability calculation are sample size, genuine performance variability across hospitals, and measurement error.

A beta-binomial statistical model, designed specifically for analyzing proportional performance measures, was utilized for reliability estimation. The beta-binomial method is specifically designed to detect and quantify measurement error (random variation) and differentiate it from true variation (systematic performance differences) between hospitals. The resulting reliability scores provide an estimate of how accurately the UNC measure reflects true differences in hospital performance.

The reliability testing involved the following steps:

1. Fitting the Beta-Binomial Model: Hospital-level UNC rates were modeled using the beta-binomial distribution.

2. Estimation of Model Parameters (alpha and beta): The beta-binomial model generates two key parameters, alpha and beta, defining the shape and variability of the underlying probability distribution, enabling calculation of variance components.

3. Calculation of Variance Components:

Between-Hospital Variance (signal): Determined from the alpha and beta parameters, reflecting the true variability in UNC performance across hospitals.

Within-Hospital Variance (noise): Calculated based on the observed proportions of UNC at each hospital, representing random measurement variability within hospitals.

4. Reliability Estimation: The ratio of between-hospital variance to the sum of between-hospital variance and within-hospital variance was calculated, resulting in a reliability score for each hospital. A higher reliability score indicates a measure with a stronger capacity to distinguish genuine performance differences among hospitals.

Missing Data:

No missing data was observed. All cases were included in the reliability analyses, and imputation methods were not necessary.

Sensitivity Analysis:

Sensitivity analysis was performed by recalculating reliability estimates after combining two years of data (2021-2022). This approach typically increases reliability by enlarging the effective sample size, reducing random measurement variability, and providing more stable estimates of hospital performance over time. The sensitivity analysis thus evaluated whether reliability improved as expected, confirming the robustness and consistency of the measure.

5.2.3 Reliability Testing Results

Reliability testing results are summarized in tables provided in section 5.2.3a. The tables include mean reliability scores by hospital deciles, as well as individual hospital performance scores and reliability statistics. For total UNC, the mean reliability score was 0.83 (Table 5.2-1), ranging from a minimum of 0.28 to a maximum of 0.98. Over 90% (179/197) of hospitals had reliability scores of 0.6 or higher, and more than 70% (140/197) achieved scores of 0.8 or above (Table 5.2-2). For severe UNC, the mean reliability score was 0.84, ranging from 0.20 to 1.00 (Table 5.2-3). Over 84% (167/197) of hospitals had reliability scores of 0.6 or higher, with 123 hospitals achieving scores of 0.8 or above (Table 5.2-4).

When combining two years of data, the reliability score improved for both total and severe UNC (Tables 5.2-5 to 5.2-8). For total UNC, only three hospitals had a reliability score below 0.6 (Table 5.2-6), while for severe UNC, the number decreased from 30 to 11 hospitals (Table 5.2-8).

5.2.3a Attach Additional Reliability Testing Results

[Suppl_Reliability.docx](#)

5.2.4 Interpretation of Reliability Results

Reliability scores range from 0.0 to 1.0. A reliability score of zero indicates that all observed variation is due to measurement error (noise), while a reliability score of 1.0 indicates that all variation reflects real differences in hospital performance. The Endorsement and Maintenance (E&M) Guidebook from Battelle recommends a reliability cutoff of 0.6 for signal-to-noise analyses.

In our reliability testing across 197 hospitals, mean reliability scores were 0.83 for total UNC and 0.84 for severe UNC, both indicating very good reliability. Mean reliability scores exceeded the acceptable threshold (0.6) in 9 out of 10 hospital deciles, with 7 to 8 deciles showing mean scores above 0.80. Approximately 9% of hospitals for total UNC and 15% for severe UNC had a reliability score below the 0.6 threshold. These lower reliability scores occurred exclusively in hospitals with smaller delivery volume (annual UNC denominators below 700 births) or those with slightly larger volumes and with outlier rates (rates above $Q3 + 1.5 \times IQR$). Small sample sizes and outliers introduce additional variability, making it challenging to differentiate true performance differences from random fluctuations or measurement errors. Additionally, reliability scores improved when two years of data were combined in a sensitivity analysis, especially for the low-volume facilities. These findings support that variations in hospital UNC scores predominantly

reflect actual differences in hospital performance, rather than measurement error. This will be discussed further in Section 5.4.1, where we compare the small hospitals without outlier values in one year versus two.

Table 2. Accountable Entity Level Reliability Testing Results by Denominator, Target Population Size

Accountable Entity-Level Reliability Testing Results													
	Overall	Minimum	Decile_1	Decile_2	Decile_3	Decile_4	Decile_5	Decile_6	Decile_7	Decile_8	Decile_9	Decile_10	Maximum
Reliability	0.830	0.280	0.514	0.689	0.789	0.834	0.869	0.888	0.911	0.932	0.949	0.964	0.980
Mean Performance Score	0.028	0.091	0.046	0.037	0.035	0.030	0.028	0.024	0.024	0.021	0.019	0.019	0.015
N of Entities	197	1	19	20	20	20	19	20	20	20	20	19	1
N of Persons / Encounters / Episodes	301,114	232	5,935	11,293	17,742	20,845	24,727	26,355	34,521	41,324	49,274	69,098	4,460

5.3.1 Level(s) of Validity Testing Conducted

Accountable entity level (i.e., measure score) (e.g., criterion validity)

5.3.2 Type of Accountable Entity Level Validity Testing Conducted

Empirical validity testing at the accountable entity-level (e.g., criterion validity, construct validity, known groups analysis)

5.3.3 Method(s) of Validity Testing

We examined several options for constructing empirical validity testing. NICU admission was not our primary choice as a comparator due to existing data demonstrating substantial variability (up to 40-fold differences) in NICU admission rates for term infants across hospitals, primarily influenced by bed availability rather than clinical necessity (Schulman et al., 2018). This variation in NICU admissions appeared to be driven by infants admitted to NICUs with transition issues.

Hospital costs adjusted from charges were utilized as a reliable indicator of morbidity, as demonstrated in recent studies on Severe Maternal Morbidity (Chen et al., 2018). Similarly, length of stay remains a validated independent indicator of illness severity (Snowden et al., 2013).

We conducted patient-level analyses to evaluate the association between Unexpected Newborn Complications and two markers of morbidity: newborn length of stay and hospital charges. We used unadjusted hospital charges directly because hospital-specific cost-to-charge ratios necessary for accurate adjustments were not available. The study population included 301,114 singleton term newborns without preexisting conditions in 197 California hospitals in 2022. Newborns who were deceased (N=25) or transferred to another facility (N=1,381) were excluded, which resulted in 299,708 newborns in this analysis. We performed descriptive univariate analysis, calculating means, standard deviations, medians, and interquartile ranges (IQR). Wilcoxon two-sample tests were used to determine if significant differences existed in the distributions of newborn length of stay and hospital charges between newborns with and without unexpected complications.

We further conducted a hospital-level analysis to assess the Pearson Correlation Coefficient between the hospital rate of unexpected newborn complications and hospital averages for newborn length of stay and hospital charges.

Reference list:

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.

5.3.4 Validity Testing Results

Validity testing results are summarized in tables provided in section 5.3.4a. The mean and median newborn length of stay and hospital charges were significantly higher among term babies with unexpected complications compared to those without complications (Table 5.3-1 & 5.3-2). Specifically, newborns with severe UNC had a mean length of stay of 5.7 days, compared to 4.7 days for those with moderate UNC and 1.7 days for those without unexpected complications. Additionally, with newborns experiencing unexpected complications, those with severe UNC consistently showed higher mean and median length of stay and hospital charges compared to those with moderate UNC.

Hospital-level analyses demonstrated positive correlations between hospital rates of unexpected newborn complications and both hospital average newborn length of stay and hospital average charges. Specifically, the correlation coefficient was 0.51 between total UNC rates and average newborn length of stay ($P < 0.01$), and 0.47 between total UNC rates and average hospital charges (Table 5.3-3).

5.3.4a Attach Additional Validity Testing Results

[Suppl_Validity.docx](#)

5.3.5 Interpretation of Validity Results

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 anticipated a correlation between UNC and indicators of newborn morbidity, specifically length of stay (LOS) and hospital charges.

The validity analyses provided strong empirical support for the UNC measure. At the patient level, newborns identified with UNC, especially severe UNC, had significantly longer hospital stays and higher hospital charges compared to newborns without complications. At the hospital level, moderate positive correlations (0.51 for length of stay and 0.47 for charges) were observed between hospital UNC rates and average resource utilization, further confirming that higher UNC rates align with increased clinical and financial burden. These consistent associations support the construct validity of UNC, indicating its utility for accurately identifying newborns with unexpected adverse outcomes.

5.4.1 Methods Used to Address Risk Factors

No risk adjustment or stratification

5.4.1b Rationale For No Adjustment or Stratification

As noted earlier, we are now recommending stratifying results by AAP NICU Levels into two categories: Level I (no NICU) and AAP Levels II-IV (hospitals with NICUs of varying intensity). The remainder of this discussion is focused on why there is not a need to risk-adjust for other factors.

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 the birth of the baby). This measure is not risk-adjusted but rather risk-stratified, using a series of exclusions (described in Section 1) 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. Newborns 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. Thus, many factors that could drive the need for risk adjustment were excluded a priori.

We did not adjust for gestational age within the term range (37-43 weeks). This decision was based on the recognition that morbidity prevalence differs across gestational ages due to varying obstetric practices, such as elective early deliveries (prior to 39 weeks) or delayed inductions beyond 41 weeks, which directly influence neonatal outcomes. The effect of these practices themselves is important to be included in the measure.

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 actors 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. Finally, our exclusions already account for most conditions typically associated with social risk factors, like preterm birth or poor fetal growth, thus no additional social risk adjustments were performed.

We paid special attention to the potential need for adjustment for race/ethnicity and payer in Section 3, Equity, and we refer you there for that discussion.

Analyses that support the decision not to adjust for patient factors are discussed below, with the conclusion that these factors did not meaningfully change hospital-level UNC rates. However, we noted variations based on hospitals' AAP neonatal care levels. AAP Level I facilities do not have a NICU and therefore have to transfer a higher proportion of infants out (which are all severe UNC), and they are typically the lowest volume facilities with reliability issues. To address these differences, we conducted stratified analyses by AAP NICU levels. Data presented subsequently will support reporting UNC rates separately for Level I hospitals and hospitals with Levels II-IV neonatal care.

ANALYSES

To investigate potential influences from unaccounted case-mix variations on our measure, we calculated risk-adjusted UNC rates at the hospital level after controlling for individual clinical and sociodemographic characteristics. We developed multivariate logistic regression models estimating the probability of severe and total UNC for each newborn. These models included maternal age, maternal pre-pregnancy BMI, payment source, prenatal care, nulliparity, infant birth weight, sex, gestational age, induction of labor, mode of delivery, and maternal comorbidities (preeclampsia, chronic hypertension, diabetes, and gestational diabetes).

Using these probabilities, we derived expected numbers of severe and total UNC events per hospital. Observed events were then compared to these expected values, resulting in an observed-to-expected (O/E) ratio. An O/E ratio greater than 1.0 indicated higher-than-expected UNC rates, while a ratio less than 1.0 indicated lower-than-expected rates. Risk-adjusted hospital UNC rates were then calculated by multiplying the O/E ratio by the population average rate (1.20% for severe UNC and 2.70% for total UNC).

Because hospital neonatal level of care emerged as the primary source of variation, we performed analysis stratifying hospitals into two groups: AAP Level I versus Levels II-IV. Within each stratum, we calculated the distribution of UNC rates and derived two performance cut-offs: the 75th percentile (Q3), designated the "alert" threshold, and the outlier threshold, defined as Q3 +

1.5 x inter-quartile range (IQR). Alert and Outlier cut points are rounded slightly to meet natural data breaks. Hospitals whose UNC rate exceeded the stratum-specific Q3 but remained below the outlier threshold were labelled "alerts". Those surpassing the outlier threshold were classified as "outliers". The lower three quartiles ($\leq Q3$) were labelled as "within expected range". Caterpillar plots were produced for each AAP level, displaying observed and adjusted rates together with the corresponding Q3 and outlier reference lines.

RESULTS AND INTERPRETATIONS

Supplementary Figure S5.4-1 displays the distribution of severe UNC rate across the 197 hospitals, and Supplementary Figure S5.4-2 contrasts observed with risk-adjusted rates in a caterpillar plot. Severe UNC rates varied widely (median 1.2%; range 0-9.1%). Risk adjustment made little difference: the mean absolute difference between observed and adjusted severe UNC rates was only 0.1 percentage points, and 85% of hospitals remained in the same quartile. This confirms that the "low risk" denominator already defines a clinically homogeneous cohort. All hospitals surpassing the high-rate threshold ($Q3 + 1.5 \times IQR \sim 3\%$) on the observed scale also exceeded that threshold after adjustment, indicating a case-mix independent performance that warrants closer monitoring. Overall, these results suggest that additional risk adjustment provides minimal added value. Comparable patterns were seen for total UNC (Supplementary Figures 5.4-3 and 5.4-4).

Supplementary Figure S.5.4-5 shows histograms of severe UNC rate, stratified by AAP NICU level.

Among Level I hospitals ($n = 43$), the distribution is right-skewed, with a median of 2.0% (IQR 1.23-2.80 %) and a long tail extending to 9.1%. Among Level II-IV hospitals ($n = 154$), rates cluster tightly around a median of 1.0 % (IQR 0.72-1.47 %), with very few hospitals above 2.5 % and a maximum of 4.6 %. The IQR in AAP Level I hospitals is roughly twice that of AAP Level II-IV hospitals. The larger variation might reflect small delivery volumes. Level I hospitals had a median annual UNC denominator < 400 births, so their rate might be inflated.

These histograms justify using AAP NICU level-specific reference lines: the 75th-percentile threshold is 2.8 % for Level I but only 1.5 % for Levels II-IV, so a single statewide cut-off would over-penalize AAP Level I hospitals and under-identify poor performers in higher-level units.

UNC is designed to help hospitals identify QI opportunities in term infants, but raw UNC rates pose some challenges for hospitals to interpret. It is a measure with low frequency, usually 1-3%, and denominator volumes can be low, especially in smaller facilities. These make small differences in rates not statistically or clinically meaningful. We are now directing hospitals to review their results and adjust their QI efforts by the category they fall into. It is difficult to establish a sharp cut point in these circumstances between normal and abnormal. Therefore, we

have chosen to create three categories based on the state-wide distribution: “Within Expected Range”; “Alert”; and “Outlier”. “Within Expected Range” accounts for all facilities in the first 3 quartiles with the QI recommendation to review Severe UNC cases selectively; “Alert” accounts for the 4th quartile excluding outliers with the recommendation for all Severe UNC cases should be routinely examined in the Perinatal QI Review Committee for improvement opportunities; lastly, “Outlier” is defined as $Q3 + 1.5 \times IQR$ beyond the start of the 4th quartile with the recommendation that there is an urgent need for formal review of Severe UNC cases including administration for potential systems improvement and additional resource requirements.

Figures S5.4-6 and S5.4-7 display caterpillar plots of observed and risk-adjusted severe UNC rates, stratified by AAP NICU level, with level-specific “alert” and “outlier” reference lines. The cut points have been adjusted slightly to make single decimal points. It is of interest that in this population, there were natural break points for the identification of “Outliers” with no category overlap, even of risk-adjusted values. Our hospitals felt there was a clear advantage to have a middle “warning” zone between “Within Expected Range” and clear “Outliers”. We recognize that the calculation of the values to use as cut points for these three categories will vary based on a given population, so they are not to be considered as fixed. An advantage of using the Box and Whiskers methodology ($Q3 + 1.5 \times IQR$) is that as hospitals improve, the number of “Outliers” can diminish to a small or even absent number, unlike relying on a fixed percentile (e.g., > 95th percentile) to establish “abnormal”.

Consistent with the plots for all hospitals, there was minimal influence of risk-adjustment on severe UNC. All hospitals with rates above the outlier thresholds remain above those thresholds after adjustment. In Level I hospitals, about one quarter of hospitals fall into the alert category, and 12% are outliers, including one extreme hospital with an adjusted rate > 8%. In Levels II-IV hospitals, about 14 % are alerts and 8% are outliers. These findings suggest that quality-improvement efforts should target hospitals above the alert levels within each NICU stratum, especially those outliers, rather than on additional risk-adjustment for hospitals performing within the expected range.

Similar findings were observed for total UNC (Supplementary Figures S5.4-8, S5.4-9, S5.4-10). In Level I hospitals, about 14% of hospitals are alert and 23% are outliers, with two extreme hospitals (adjusted rates > 9%). In Levels II-IV hospitals, about 16 % are alerts and 6% are outliers. The consistency across severe and total UNC further supports directing QI efforts toward the relatively small subset of hospitals with persistently elevated rates.

For the smallest of hospitals with deliveries <400 term births per year, we were concerned about reliability, especially when there were high rates. To address this question, we compared the distribution of AAP Level I hospitals over a two-year period versus a single year. Supplementary Figures S5.4-11 and S5.4-12 illustrate observed and risk-adjusted severe and total UNC rates for

53 AAP Level I hospitals after pooling 2021 and 2022 data. Aggregating two years did not change hospital ordering or risk-adjusted values. Importantly, 4 of 5 severe UNC outliers and 6 of 9 total UNC outliers identified in 2022 remained outliers in the two-year analysis, demonstrating that elevated rates at these facilities are persistent rather than year-to-year noise. These findings underscore that even with larger denominators and despite risk adjustment, performance gaps exist. As this is not a publicly reported measure, we feel quite comfortable identifying “Outliers” using a single year’s data to help drive the hospital’s QI efforts.

6.1.1 Current Status

In use

6.1.3 Current Use(s)

Quality Improvement with Benchmarking (external benchmarking to multiple organizations),
Quality Improvement (Internal to the specific organization)

6.1.3 Program Details

Name of the program and sponsor

The Joint Commission Perinatal Care Measures

URL of the program

<https://www.jointcommission.org/measurement/measures/perinatal-care/>

Purpose of the program

The Joint Commission Perinatal Care measures focus on achieving integrated, coordinated, patient-centered care for clinically uncomplicated pregnancies and births.

Geographic area and percentage of accountable entities and patients included

The Joint Commission accredits almost 80% of U.S. hospitals, covering over 4,000 hospitals nationwide.

Applicable level of analysis and care setting

Hospital

,

Name of the program and sponsor

California Maternal Quality Care Collaborative

URL of the program

<https://www.cmqcc.org/>

Purpose of the program

The California Maternal Quality Care Collaborative (CMQCC) is a multi-stakeholder organization committed to ending preventable morbidity, mortality and racial disparities in maternity care across California, Oregon, and Washington.

Geographic area and percentage of accountable entities and patients included

CMQCC covers over 99% of hospitals and patients in California and more than 75% of delivery volume in Oregon and Washington.

Applicable level of analysis and care setting

Hospital

6.2.1 Actions of Measured Entities to Improve Performance

The morbidity assessed in this measure represents the collective effect of BOTH obstetric and neonatal care. As a result, both services may have QI opportunities identified by higher rates. In turn, we have assisted facilities in successful QI projects in both obstetrics and neonatology. The specific topics are listed below.

- 1) Management of infants at risk for early onset sepsis: This is the most common neonatal issue that leads to elevated rates. Common issues have been the use of C-reactive Protein to screen for sepsis (in 6 California hospitals after CRP was replaced by the Kaiser EOS Calculator, their Severe UNC rates fell by an average of 50%).
- 2) Safe use of vacuum extractor at delivery: Vacuum extractors, if not used with an appropriate protocol, can lead to a variety of neonatal head injuries. These are measured in UNC and can be benchmarked. Hospitals have used this information to review their Vacuum protocols and reduce head injuries.
- 3) As cesarean rates have come under increasing scrutiny, hospitals have been very interested in a balancing measure of term neonatal health as they work to reduce cesarean rates. This has been discussed at length elsewhere.
- 4) The large variation of severe UNC among small hospitals is almost exclusively driven by transfers to higher levels of care, usually among infants with transition problems. Most small hospitals have trained Family Medicine physicians and nurses to better support these infants for their first 2-3 hours of life without the need to separate the infant from the mother in a transfer to a distant higher-level facility. As other small hospitals see the success of their peer hospitals, these practices are spreading. The State Perinatal Quality Collaborative has seen this as an opportunity to support the sharing of best practices for term infants. The use of comparison metrics, such as UNC, is a key driver to incentivize hospitals to change practices.

6.2.2 Feedback on Measure Performance

CMQCC Maternal Data Center solicits regular feedback on the quality measures that we support from hospitals in California, Oregon, and Washington. Specifically, we have several webinars per year on UNC, including how to interpret it and use it for hospital QI. These have been a rich source of feedback. We receive several questions every month regarding specific coding and implementation questions. These have been very useful as we update ICD-10 codes every year. In addition, we regularly share feedback with the Joint Commission, which uses this measure as PC-06 and receives feedback from an even larger set of hospitals. The most important feedback has been from small rural hospitals regarding the issue around transports. As discussed above, most small rural hospitals do well, but as they do not have a NICU, they can have higher rates due to the need for transfer of some infants. This is a key reason that we are now proposing a separate category for AAP Level I facilities so they can compare themselves to each other. We have done this in California for the last 3 years and it has been very instructive for all.

6.2.3 Consideration of Measure Feedback

Feedback has led to multiple measure improvements over the last several years. A major improvement was to add LOS modifiers to guard against over-coding and to move a few diagnoses from severe to moderate after feedback. Every year, we get code ideas (and questions) from our users. The biggest recommendation from users during this review period is to benchmark AAP Level I hospitals separately.

6.2.4 Progress on Improvement

We have demonstrated substantial statewide improvement. Our recent trend analysis from 237 California hospitals between 2016 and 2022 shows a notable statewide relative rate reduction of approximately 20% in severe UNC, decreasing from 1.50% in 2016 to 1.20% in 2022. Total UNC also experienced an overall reduction of 7.64% (Supplementary Table S6.1). The current UNC definition (and codes) was used when comparing prior years to the current.

Our prior publications further demonstrate the effectiveness of using severe UNC as a balancing measure in large-scale cesarean reduction programs. Main (2019) showed that a state-wide QI initiative successfully reduced the nulliparous, term, singleton, vertex (NTSV) cesarean delivery rate from 29.3% in 2015 to 25.0% in 2017 without increasing the severe UNC rate (it actually was reduced). Similarly, Rosenstein (2021) further confirmed significant statewide reductions in cesarean rates without compromising neonatal outcomes as measured by severe UNC, further supporting the utility of severe UNC as a reliable balancing measure.

At the hospital level, we have observed meaningful improvements in hospitals actively addressing their UNC rates, achieving reductions between 10% to 20%. Currently, no formal statewide QI collaborative specifically targeting this measure. Instead, the measure continues to serve effectively as a balancing measure within broader improving vaginal birth QI activities, ensuring that improvements in maternal outcomes do not negatively impact neonatal health. In one substantial pilot involving three hospitals and over 10,000 births over two years, a 20% reduction in cesarean deliveries was achieved without any increase in adverse newborn outcomes (measured by UNC, Fineberg 2024).

Collectively, these studies and experiences described in 6.2.1 underscore the effectiveness and importance of UNC as a robust balancing measure, safeguarding neonatal outcomes during maternal-focused quality improvement interventions.

Reference list:

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.

Rosenstein MG, Chang SC, Sakowski C, Markow C, Teleki S, Lang L, Logan J, Cape V, Main EK. Hospital Quality Improvement Interventions, Statewide Policy Initiatives, and Rates of Cesarean Delivery for Nulliparous, Term, Singleton, Vertex Births in California. JAMA. 2021 Apr 27;325(16):1631-1639.

Fineberg AE, Harley K, Lahiff M, Main EK. The relative impact of labor induction versus improved labor management: Before and after the ARRIVE (a randomized trial of induction vs. expectant management) trial. Birth. 2024 Dec;51(4):719-727. doi: 10.1111/birt.12845.

6.2.5 Unexpected Findings

No unintended findings have been noted. It has been successful as a tool for improving care in individual hospital QI projects and for larger system and state-wide improvement initiatives. We also encourage hospitals to use this measure as a balancing measure to identify unintended consequences of other National QI measures, such as Cesarean Sections.

7.1 Supplemental Attachment

[supplementary information.docx](#)

Developer POC email

dlyell@stanford.edu

Measure Developer POC

Deirdre Lyell
California Maternal Quality Care Collaborative
Palo Alto, CA
United States

The measure developer is different from the measure steward

No

Steward Address

Deirdre Lyell
Palo Alto, CA
United States

Steward Organization

California Maternal Quality Care Collaborative

Steward Organization URL

<https://www.cmqcc.org/>

Steward POC email

dlyell@stanford.edu