

# Perinatal Risk Factors and Outcome Coding in Clinical and Administrative Databases

Daniel S. Tawfik, MD, MS,<sup>a,b</sup> Jeffrey B. Gould, MD, MPH,<sup>c,d</sup> Jochen Profit, MD, MPH<sup>c,d</sup>

abstract

**BACKGROUND AND OBJECTIVES:** Administrative databases may allow true population-based studies and quality improvement endeavors, but the accuracy of billing codes for capturing key risk factors and outcomes needs to be assessed. We sought to describe the performance of a statewide administrative database and the clinical database from the California Perinatal Quality Care Collaborative (CPQCC).

**METHODS:** This population-based retrospective cohort study linked key perinatal risk factors and outcomes from the 133-unit CPQCC database to relevant billing codes from administrative maternal and newborn inpatient discharge records, for 50 631 infants born from 2006 to 2012. Using the CPQCC record as the gold standard, we calculated the positive predictive value, negative predictive value, and Matthews correlation coefficient for each item, then evaluated comparative performance across units.

**RESULTS:** The Matthews correlation coefficient was highest (>0.7; strong positive correlation) for multiple delivery, Cesarean delivery, very low birth weight, maternal hypertension, maternal diabetes, patent ductus arteriosus, in-hospital death, patent ductus arteriosus and retinopathy of prematurity surgeries, extracorporeal life support, and intraventricular hemorrhage. Maternal chorioamnionitis, fetal distress, retinopathy of prematurity staging, chronic lung disease, and pneumothorax were the least reliably coded. Maternal factors and delivery details were more reliably coded in the maternal inpatient record than the newborn inpatient record.

**CONCLUSIONS:** Several important perinatal risk factors and outcomes are highly congruent between these administrative and clinical databases. Several subjective risk factors and outcomes are appropriate targets for data improvement initiatives. The ability for timely extraction of administrative inpatient data will be key to their usefulness in quality metrics.



Divisions of <sup>a</sup>Perinatal Epidemiology and Health Outcomes Research Unit, Neonatology and Departments of <sup>a</sup>Pediatric Critical Care Medicine, Pediatrics and <sup>b</sup>Health Research and Policy, School of Medicine, Stanford University, Palo Alto, California; and <sup>c,d</sup>California Perinatal Quality Care Collaborative, Palo Alto, California

Dr Tawfik conceptualized and designed the study, conducted the analyses, drafted the initial manuscript, and reviewed and revised the manuscript; Drs Gould and Profit conceptualized and designed the study, coordinated and supervised data analysis, and critically reviewed the manuscript for important intellectual content; and all authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

**DOI:** <https://doi.org/10.1542/peds.2018-1487>

Accepted for publication Sep 14, 2018

Address correspondence to Daniel S. Tawfik, MD, MS, Division of Pediatric Critical Care Medicine, Department of Pediatrics, Stanford University School of Medicine, 770 Welch Rd, Suite 435, Stanford, CA 94304. E-mail: dtawfik@stanford.edu

**WHAT'S KNOWN ON THIS SUBJECT:** Administrative databases containing diagnosis and procedure codes from discharge records hold potential for population-based studies and quality improvement endeavors. However, the accuracy of coding practices for these purposes is poorly understood, and targets for data quality improvement efforts are unknown.

**WHAT THIS STUDY ADDS:** Through linkage of a statewide administrative database to a large clinical quality care collaborative, we found high accuracy of administrative coding for common and well-defined perinatal risk factors and outcomes. However, several other items represent data quality improvement opportunities.

**To cite:** Tawfik DS, Gould JB, Profit J. Perinatal Risk Factors and Outcome Coding in Clinical and Administrative Databases. *Pediatrics*. 2019;143(2):e20181487

Quality benchmarking, value-based payment structures, and population-based studies all require accurate and efficient data collection to be successful. Administrative databases are appealing to use for these purposes, with rich data readily available through the ongoing inclusion of discharge diagnosis and procedure codes.<sup>1,2</sup> However, the accuracy of these coding practices is poorly understood and may be subject to overdiagnosis, omissions, or misclassifications depending on local coding practices and incentives.<sup>3–5</sup> The primary purposes of such databases are financial and administrative rather than clinical, and thus discrepancies are likely to occur particularly for clinical conditions without strong administrative or financial implications.<sup>6</sup> In contrast, clinical databases contain high-quality comprehensive data but are labor intensive and expensive to maintain.

The California Perinatal Quality Care Collaborative (CPQCC) maintains a robust and high-quality clinical database based on newborn stays at all California Children's Services–accredited NICUs in California.<sup>7</sup> In parallel to CPQCC, all NICUs in California also report administrative data to the Office of Statewide Health Planning and Development (OSHPD), including *International Classification of Diseases* diagnosis and procedure codes for all patients. Using this overlap of clinical and administrative databases on a statewide level, we sought to describe the performance of the administrative inpatient discharge database from OSHPD compared with the clinical database from CPQCC, for infant and maternal risk factors and infant outcomes.

## METHODS

The California OSHPD provides a linked vital statistics birth, newborn

discharge, and maternal delivery data file. Data for these files are submitted directly to OSHPD from designated hospital staff or a designated reporting agent at minimum twice yearly. Data for the CPQCC database are submitted directly to CPQCC in real time from bedside nurses and dedicated data abstractors who undergo yearly training, with logic and range checks at the time of data entry, as well as confirmation when records exceed defined thresholds for missing or unobtainable items. In addition, consistency checks are employed for infants who transfer hospitals. We conducted a probabilistic record linkage of the CPQCC database to the OSHPD maternal and infant files for 2006–2012 on the basis of infant date of birth, maternal date of birth, infant sex, birth weight, birth location, infant disposition, infant discharge date, and birth order.

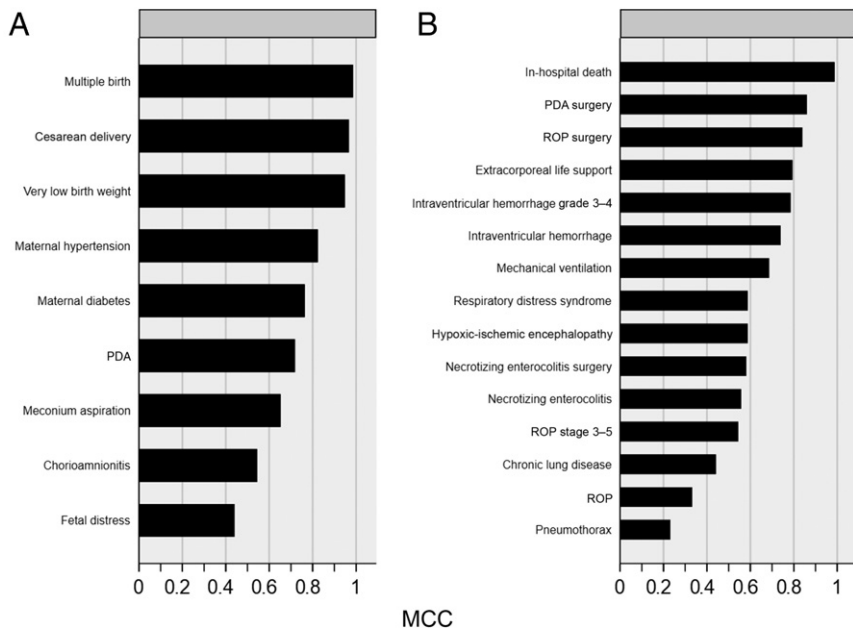
We included all inborn infants who were born in 2006–2012, met eligibility criteria at a CPQCC NICU, and who were discharged from the hospital or died. CPQCC eligibility criteria include all infants with a birth weight between 401 and 1500 g or gestational age between 22 + 0/7 and 29 + 6/7 weeks' gestation, in addition to all infants admitted before 28 days of age and meeting any of the following criteria: death before discharge, acute transfer into an NICU, acute transfer out of an NICU, major surgery requiring anesthesia, assisted ventilation for >4 hours, nasal intermittent mandatory ventilation for >4 hours, early bacterial sepsis, readmission for total serum bilirubin >25 mg/dL, or exchange transfusion.

We extracted perinatal risk factor and outcomes data from this CPQCC record. Perinatal risk factors included birth weight, Cesarean delivery, multiple delivery, fetal distress, meconium aspiration, and

patent ductus arteriosus (PDA), as well as maternal chorioamnionitis, hypertension (pregestational and gestational, including preeclampsia), and diabetes (pregestational and gestational). Infant outcomes included mechanical ventilation, pneumothorax, respiratory distress syndrome, chronic lung disease, necrotizing enterocolitis, intraventricular hemorrhage, hypoxic-ischemic encephalopathy, retinopathy of prematurity (ROP), extracorporeal life support, and in-hospital death. In addition, we evaluated combined grade 3 and grade 4 intraventricular hemorrhage and combined stages 3 through 5 ROP to align with common categorization of these more clinically important outcomes.

We then added relevant *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM) diagnosis and *International Classification of Diseases, Ninth Revision, Procedure Coding System* (ICD-9-PCS) procedure codes from the OSHPD maternal and newborn inpatient discharge records, as outlined in Supplemental Table 3. For the primary analysis, *International Classification of Diseases, Ninth Revision* codes in either the maternal or infant record were considered to be coded in the administrative database. Using the CPQCC record as the gold standard, we then developed 2 × 2 tables and calculated several statistical measures for each risk factor and outcome.

Because of the highly variable group sizes, particularly for rare events, our primary measure of interest was the Matthews<sup>8</sup> correlation coefficient (MCC) as a summary estimate of the quality of classification.<sup>9</sup> MCC ranges from –1 (complete disagreement) to 1 (complete agreement), with a value of 0 representing predictive ability



**FIGURE 1** MCCs for risk factor and outcome coding, compared between an administrative and clinical database. A, Risk factors. B, Outcomes.

equal to random chance. Because MCC is a specific application of the Pearson correlation coefficient, we interpreted the strengths of associations accordingly, with 0.1 to 0.4 indicating weak positive correlation, 0.4 to 0.7 indicating moderate positive correlation, and 0.7 to 1 indicating strong positive correlation. Our secondary measures of interest included the positive predictive value (PPV) (proportion of positives that are true-positives), because this provides an intuitive measure of the precision of the test, or the ability to avoid false-positives. In a similar way, we calculated the negative predictive value (NPV) (proportion of negatives that are true-negatives) as a secondary measure of interest.

To evaluate the interhospital variability, we calculated summary statistics of MCC performance on all risk factors and outcomes, aggregated for each hospital in the sample. We then compared organizational factors between the highest-performing 10 hospitals to the lowest-performing 10 hospitals via Wilcoxon rank test for

continuous variables and Fisher's exact test for categorical variables. All tests were 2 sided with an  $\alpha$ -level of .05.

Approval for this study was obtained through the Stanford University Institutional Review Board with a waiver of informed consent. Data analysis was performed using SAS (SAS Institute, Inc, Cary, NC), version 9.4, and Stata (Stata Corp, College Station, TX), version 15.0.

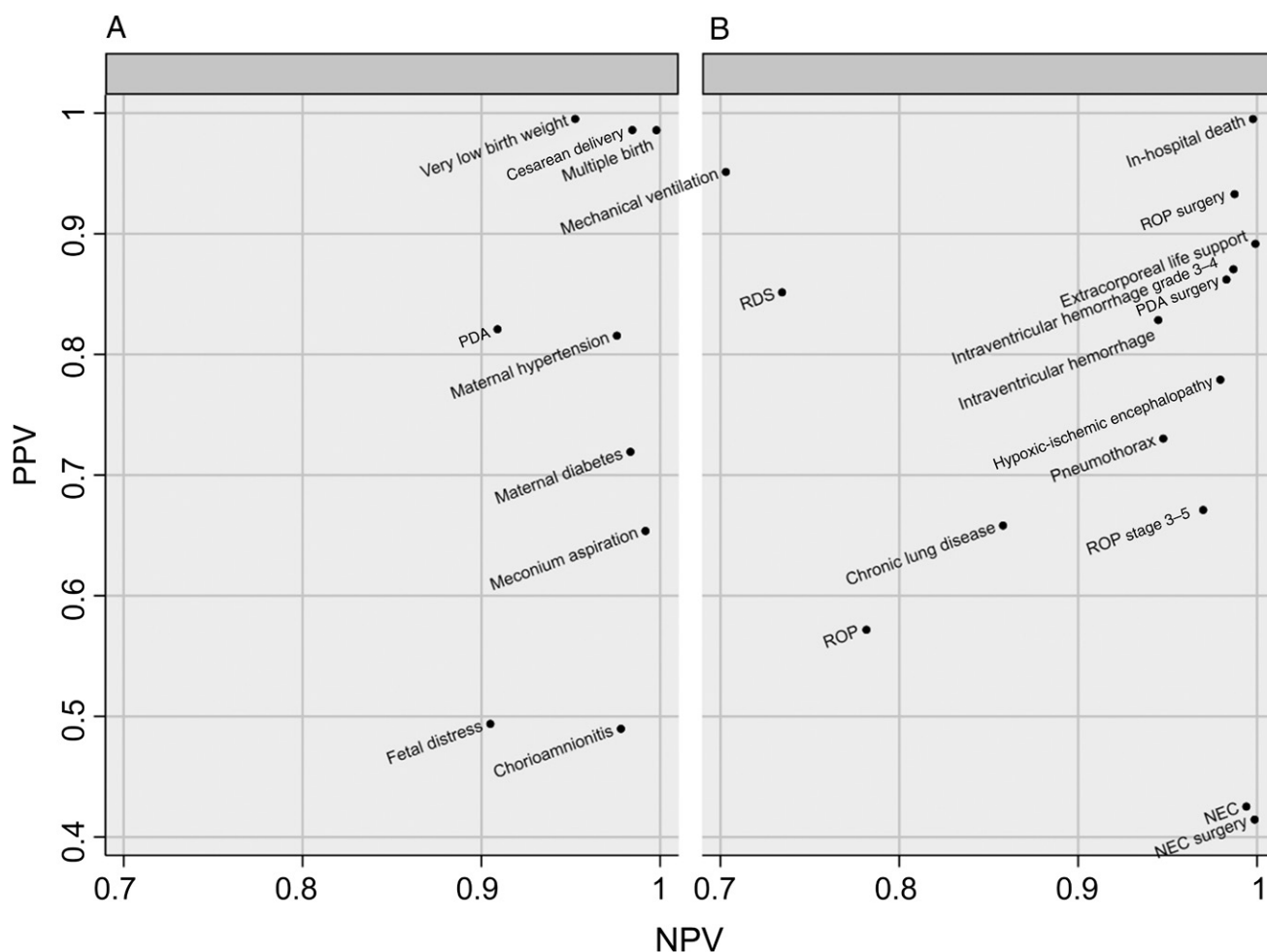
## RESULTS

A total of 50 631 newborns were successfully linked out of 51 612 eligible records, representing a 98.1% linkage rate. Ranking of risk factors and outcomes by MCC value is shown in Fig 1, with graphical representation of PPV in relation to NPV shown in Fig 2.

MCC was  $>0$  (positive correlation) for all risk factors, with an MCC  $>0.9$  (very strong positive correlation) and PPV  $>90\%$  for the risk factors of very low birth weight, cesarean delivery, and multiple-gestation delivery,

as shown in Table 1. Similarly, an MCC  $>0.7$  and PPV of 70% to 90% were observed for the risk factors of maternal hypertension, maternal diabetes, and PDA. MCC was  $<0.7$  with a PPV  $<70\%$  for the remainder of the risk factors evaluated. NPV was  $>90\%$  for all risk factors. All variables other than in-hospital death revealed notable variation in PPV among hospitals as shown in Table 1. In particular, extracorporeal life support, pneumothorax, intraventricular hemorrhage, PDA surgery, and hypoxic-ischemic encephalopathy each exhibited a PPV of 100% in  $>25\%$  of hospitals despite overall PPV values of  $<90\%$  among the sample.

We also calculated  $2 \times 2$  tables separately for risk factors coded in the maternal OSHPD record and the infant OSHPD record, as shown in Table 2. Cesarean delivery and multiple delivery revealed good precision, with a PPV of 98% or higher in both maternal and infant records. However, MCC and NPV were lower in the infant record compared with maternal record for all variables. Because a coded risk factor in either the maternal record or the infant record was considered to be a case as reported in Table 1, it was more common to have negative administrative items in the divided risk factors as reported in Table 2. For example, maternal diabetes was negative in 43 595 of maternal records and in 47 845 of infant records, but because of incomplete overlap, it was only negative in 43 244 of the combined records. This corresponds to a smaller number of true-negatives when compared with the 44 468 negative records from the CPQCC database. Similar trends were seen for hypertension and fetal distress. For cesarean delivery and multiple delivery, the maternal record captured all of the cases in the infant record, as well as a few additional cases, so no similar discrepancy existed for these items.



**FIGURE 2** The PPV in relation to the NPV for risk factor and outcome coding, compared between an administrative and clinical database. A, Risk factors. B, Outcomes. NEC, necrotizing enterocolitis; RDS, respiratory distress syndrome.

For patient outcomes, MCC was >0 for all, with MCC >0.9 only for in-hospital death. MCC >0.7 and a PPV of 70% to 90% were observed for PDA surgery, ROP surgery, extracorporeal life support, and intraventricular hemorrhage. MCC was <0.7 for the remainder of the outcomes evaluated, with MCC <0.4 for ROP and pneumothorax. NPV was >90% for all outcomes except chronic lung disease (86%), ROP (78%), respiratory distress syndrome (73%), and mechanical ventilation (70%).

Interhospital variability in performance as evaluated by mean MCC revealed similar performance among the majority

of hospitals, with deviation at the extremes, as shown in Fig 3. There were no significant differences between the 10 highest-performing hospitals and the 10 lowest-performing hospitals, when comparing organizational factors encompassing patient volume, acuity, sociodemographics, and ownership (see Supplemental Table 4).

### DISCUSSION

With this large, population-based study of NICU patients, we identified successes and opportunities for improvement in the use of administrative billing codes for perinatal risk factors and outcomes. Although many of the items

evaluated performed well (nearly half of the MCCs were >0.7), we observed variation in performance across the items.

Risk factors and outcomes that are highly prevalent or easy to define revealed the most reliable coding in the OSHPD database. This finding is in line with those of Ford et al,<sup>10</sup> who found particularly high accuracy of procedural coding in a similar study of 2432 infants but with more subjective diagnoses such as transient tachypnea performing poorly. With our findings, we suggest that usage of administrative data for very low birth weight designation, cesarean delivery, or multiple delivery determinations allows for

**TABLE 1** Comparison of Perinatal Risk Factor and Outcome Coding, Between an Administrative and Clinical Database

Risk factors	N <sup>a</sup>	True-Negatives, n (%)	False-Negatives, n (%)	False-Positives, n (%)	True-Positives, n (%)	PPV, % <sup>b</sup> (IQR) <sup>c</sup>	NPV <sup>d</sup> %	MCC
Very low birth wt	50631	24 127 (47.7)	1206 (2.4)	122 (0.2)	25 176 (49.7)	100 (99–100)	95	0.95
Cesarean delivery	50618	16 490 (32.6)	262 (0.5)	475 (0.9)	33 391 (66.0)	99 (97–100)	98	0.97
Multiple delivery	50607	39 610 (78.3)	86 (0.2)	154 (0.3)	10 757 (21.3)	99 (99–100)	100	0.99
Meconium aspiration	50631	48 963 (96.7)	409 (0.8)	436 (0.9)	823 (1.6)	65 (50–83)	99	0.65
Fetal distress	50208	33 356 (66.4)	3502 (7.0)	6758 (13.5)	6592 (13.1)	49 (35–63)	90	0.44
Maternal hypertension	50177	37 276 (74.3)	927 (1.8)	2209 (4.4)	9765 (19.5)	82 (74–90)	98	0.82
Maternal diabetes	50176	42 522 (84.7)	722 (1.4)	1946 (3.9)	4986 (9.9)	72 (54–84)	98	0.76
PDA	43607	28 490 (65.3)	2852 (6.5)	2197 (5.0)	10 068 (23.1)	82 (73–90)	91	0.72
Maternal chorioamnionitis <sup>e</sup>	35728	31 893 (89.3)	717 (2.0)	1591 (4.5)	1527 (4.3)	49 (17–57)	98	0.54
Outcomes								
Mechanical ventilation	50631	15 888 (31.4)	6712 (13.3)	1366 (2.7)	26 665 (52.7)	95 (92–98)	70	0.69
Respiratory distress syndrome	50631	19 290 (38.1)	6975 (13.8)	3620 (7.1)	20 746 (41.0)	85 (77–93)	73	0.59
In-hospital death	50554	44 518 (88.1)	102 (0.2)	29 (0.1)	5905 (11.7)	100 (100–100)	100	0.99
Extracorporeal life support	48262	48 098 (99.7)	44 (0.1)	13 (0.0)	107 (0.2)	89 (89–100)	100	0.79
Pneumothorax	48245	45 427 (94.2)	2514 (5.2)	82 (0.2)	222 (0.5)	73 (67–100)	95	0.23
Necrotizing enterocolitis	48228	45 798 (95.0)	276 (0.6)	1238 (2.6)	916 (1.9)	42 (20–54)	99	0.56
Necrotizing enterocolitis surgery	48225	47 430 (98.4)	64 (0.1)	428 (0.9)	303 (0.6)	41 (15–53)	100	0.58
Intraventricular hemorrhage <sup>f</sup>	34392	27 325 (79.5)	1598 (4.6)	938 (2.7)	4531 (13.2)	83 (74–92)	94	0.74
Grade 3–4 intraventricular hemorrhage <sup>f</sup>	34392	32 641 (94.9)	437 (1.3)	170 (0.5)	1144 (3.3)	87 (80–100)	99	0.78
Chronic lung disease <sup>g</sup>	22769	16 694 (73.3)	2763 (12.1)	1132 (5.0)	2180 (9.6)	66 (50–80)	86	0.44
ROP <sup>h</sup>	19345	11 347 (58.7)	3171 (16.4)	2067 (10.7)	2760 (14.3)	57 (39–73)	78	0.33
Stage 3–5 ROP <sup>h</sup>	19345	18 028 (93.2)	560 (2.9)	249 (1.3)	508 (2.6)	67 (33–88)	97	0.54
ROP surgery <sup>h</sup>	19298	18 247 (94.6)	232 (1.2)	55 (0.3)	764 (4.0)	93 (93–100)	99	0.84
PDA surgery <sup>i</sup>	12914	10 783 (83.5)	188 (1.5)	288 (2.1)	1675 (13.0)	86 (86–100)	98	0.86
Hypoxic-ischemic encephalopathy <sup>j</sup>	12145	11 634 (95.8)	244 (2.0)	59 (0.5)	208 (1.7)	78 (62–100)	98	0.59

IQR, interquartile range.

<sup>a</sup> N = 50631 infants born at a CPQCC NICU with disposition home or death with linked clinical and administrative records. Clinical database values are used as the gold standard for comparisons.

<sup>b</sup> PPV = (true-positives/all positives) × 100%.

<sup>c</sup> IQR reflecting performance of individual hospitals.

<sup>d</sup> NPV = (true-negatives/all negatives) × 100%.

<sup>e</sup> Collected in CPQCC starting from 2008.

<sup>f</sup> For infants with cranial image by day of life 28.

<sup>g</sup> Defined as supplemental oxygen at 36 wk corrected gestational age.

<sup>h</sup> For infants with very low birth wt with eye examination.

<sup>i</sup> For infants diagnosed with PDA.

<sup>j</sup> For infants born at 36 wk completed gestation or later.

**TABLE 2** Comparison of Perinatal Risk Factor Coding for Maternal and Infant Administrative Records, Compared With a Clinical Database

	ICD-9-CM and ICD-9-PCS Codes	N <sup>a</sup>	True-Negatives, n (%)	False-Negatives, n (%)	False-Positives, n (%)	True-Positives, n (%)	NPV, <sup>b</sup> %	PPV, <sup>c</sup> %	MCC
Cesarean delivery									
Maternal record	669.7x, 74.0, 74.1, 74.2, 74.4, 74.5, 74.9x	50618	16490 (32.6)	262 (0.5)	475 (0.9)	33391 (66.0)	98	99	0.97
Infant record	V3x01	50618	16449 (32.5)	981 (1.9)	516 (1.0)	32672 (64.5)	94	98	0.93
Multiple delivery									
Maternal record	651.x, V27.x, V27.3, V27.4, V27.5, V27.6, V27.7	50607	39610 (78.3)	86 (0.2)	154 (0.3)	10757 (21.3)	100	99	0.99
Infant record	761.5, V31.x, V32.x, V33.x, V34.x, V35.x, V36.x, V37.x	50607	39653 (78.4)	347 (0.7)	111 (0.2)	10496 (20.7)	99	99	0.97
Fetal distress									
Maternal record	655.7x, 656.3x, 659.7x	50208	33510 (66.7)	3620 (7.2)	6604 (13.2)	6474 (12.9)	90	50	0.44
Infant record	763.82, 768.2, 768.3, 768.4	50208	39744 (79.2)	9406 (18.7)	370 (0.7)	688 (1.4)	81	65	0.16
Maternal hypertension									
Maternal record	642.x, 401.x, 402.x, 403.x, 404.x, 437.2	50177	37310 (74.4)	968 (1.9)	2175 (4.3)	9724 (19.4)	97	82	0.82
Infant record	760.0	50177	39281 (78.3)	9129 (18.2)	204 (0.4)	1563 (3.1)	81	88	0.31
Maternal diabetes									
Maternal record	648.0, 648.8, 790.2, 250.x, 249.x	50176	42696 (85.1)	899 (1.8)	1772 (3.5)	4809 (9.6)	98	73	0.75
Infant record	775.0	50176	43992 (87.7)	3853 (7.7)	476 (0.9)	1855 (3.7)	92	80	0.47
Maternal chorioamnionitis <sup>d</sup>									
Maternal record	658.4x, 659.21, 659.31	35728	32001 (89.6)	788 (2.2)	1483 (4.2)	1456 (4.1)	98	50	0.53
Infant record	762.7	35728	33309 (93.2)	1861 (5.2)	175 (0.5)	383 (1.1)	95	69	0.32

<sup>a</sup> N = 50631 infants born at a CPQCC NICU with disposition home or death with linked clinical and administrative records. Clinical database values are used as the gold standard for comparisons.

<sup>b</sup> NPV = (true-negatives/all negatives) × 100%.

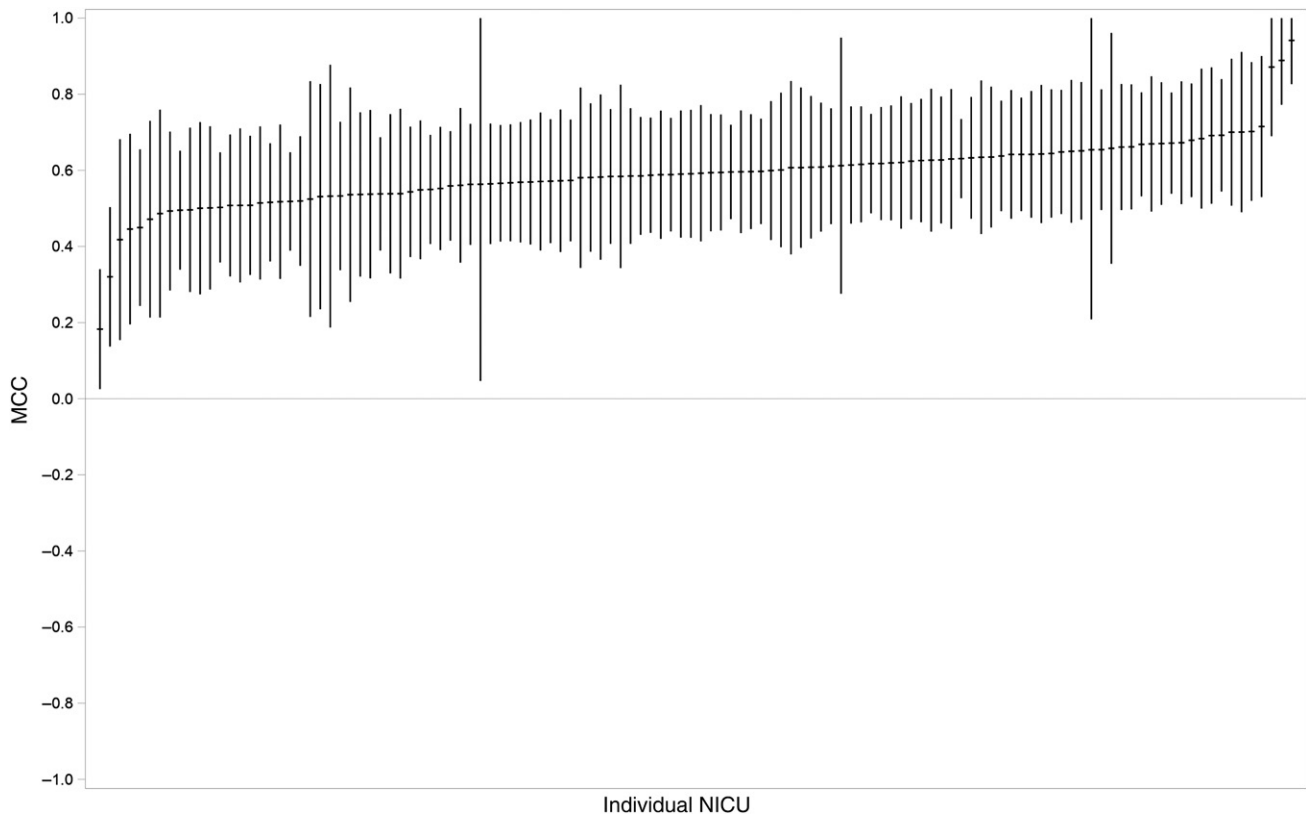
<sup>c</sup> PPV = (true-positives/all positives) × 100%.

<sup>d</sup> Collected in CPQCC starting from 2008.

similar results as usage of a clinical database. Similarly, mortality, mechanical ventilation, and ROP surgery remain highly congruent between the 2 types of databases. Conversely, diagnoses with subjective or complicated definitions, such as maternal chorioamnionitis, fetal distress, and chronic lung disease, performed less well. These entities that are more challenging to define are likely to represent high-yield targets of data improvement initiatives or incentive realignment.

Among infant risk factors, usage of a combination of the maternal and infant records reduced the false-negative rate over use of the infant record alone. This discrepancy was particularly pronounced for maternal factors that may not be as readily recognizable to neonatal providers, such as maternal chorioamnionitis, maternal hypertension, and maternal diabetes. It is suggested with these results that additional caution is needed when interpreting prevalence of maternal factors using infant records alone; with these results, the opportunity that arises when multiple records feature overlapping target areas is also highlighted. With increasingly ubiquitous electronic health records, it is conceivable that automated duplication of relevant codes between maternal and infant health records could be used to minimize discrepancies.

Of note, even the highest-performing metrics exhibited some degree of discrepancy between the 2 databases, indicating that all factors have some room for improvement. Very low birth weight, cesarean delivery, and multiple delivery all represent readily identifiable risk factors but featured hundreds of patients in each group with discrepant coding. Similarly, discrete and high-profile outcomes such as extracorporeal life support and mortality did not exhibit complete congruence. The reasons for these discrepancies



**FIGURE 3** Comparative performance between an administrative and clinical database coding across hospitals. Mean MCC with 95% confidence intervals are shown for each hospital.

are unclear, but their existence can be used to highlight the potential for improvement across the coding spectrum. This potential for improvement is particularly relevant for rare occurrences such as extracorporeal life support, because small numbers of discrepancies can represent a large proportion of rare cases. Additionally, although the degree of discrepancy among the more common high-performing factors is unlikely to materially affect analyses or conclusions at the population level, they can contribute to multiplicative effects when multiple factors are evaluated simultaneously.

Researchers of previous evaluations of administrative claims records have found them to exhibit use for identification of rare cases or clinical areas for improvement,<sup>2,4,11</sup> but less reliable for comparative

evaluations or identification of absolute incidence.<sup>3,12,13</sup> In addition, ancillary financial data such as insurance coverage have been found to have poor reliability.<sup>14</sup> However, unique characteristics of the NICU population lend themselves well to the usage of administrative claims data, including linkage between infant and maternal records and readily quantifiable risk factors such as birth weight and mode of delivery. This concept was recently illustrated by Howell et al<sup>15</sup> in their analysis of maternal sociodemographic characteristics in relation to neonatal morbidity and mortality. Usage of linked maternal and infant data can also allow for increased power in the evaluation of population-level risk factors and expand horizons for health services research.

Similarly, comprehensive and accurate data are required for

adequate performance metrics and quality improvement initiatives. Often, these goals are achieved through the development of clinical databases as part of quality care collaboratives. To be effective, these collaboratives must develop accurate methods of data collection, then seek to expand and become more comprehensive, as demonstrated by the CPQCC used in this study.<sup>7</sup> However, many states and regulatory agencies lack the resources, regulatory mechanisms, or provider leadership to establish large clinical databases in this way, and even those with sufficient resources are unlikely to accomplish such a feat rapidly. With our findings, we suggest that an alternative approach is feasible, in which stakeholders instead seek to improve the accuracy of administrative data, leveraging the already-comprehensive nature of these databases. This alternative

administrative-based approach could be used to support statewide quality improvement efforts such as the National Network of State Perinatal Quality Collaboratives supported currently by the Centers for Disease Control and Prevention.<sup>16</sup>

Compared with the creation of a novel quality care collaborative, improving the quality of data in existing administrative databases could be accomplished relatively efficiently through investigation, feedback, education, and incentive realignment. Data quality improvement should begin with a thorough investigation into the contributory factors to inaccurate coding, with the development of targeted improvement measures accordingly. Models already exist for feedback and education, including feeding back clinically relevant analyses of administrative data to clinicians and establishing regular data quality improvement seminars for medical coders.<sup>7,17</sup> Use of these data for performance metrics would also likely provide a strong incentive to accelerate data quality improvement efforts locally. If coupled with an implementation arm to effect change, the existing infrastructure may thus be leveraged to simultaneously improve outcomes for patients, increase efficiency for quality improvement programs, and reduce costs for public and private payers.

Although we observed interhospital variation among summary performance metrics, the magnitude of this variation was relatively small for the majority of the hospitals in this study. This suggests that current incentives are insufficient to promote an emphasis on high-quality administrative data collection at the majority of these sites. Implementing widespread data quality improvement efforts will thus be particularly important, because the results do not appear to be limited to any particular subset of hospitals.

This study must be interpreted in the context of its design. The probabilistic matching process carries the risk of false matches, which would be expected to bias the results toward lower PPVs. However, the large number of demographic variables shared between the 2 databases minimizes this possibility, and this matching process has shown good reliability in previous evaluations.<sup>18</sup> Our usage of the CPQCC database as the gold standard for purposes of analysis reflects our objective to compare the administrative claims database to the currently available best alternative. Although CPQCC is likely not 100% accurate in coding, it employs extensive quality control measures and represents the most accurate currently available alternative to the usage of administrative claims records for this population, with definitions matching those used by the Vermont Oxford Network.<sup>19,20</sup> The ICD-9-CM and ICD-9-PCS codes used for case identification aimed for consistency with their usage in the current literature, but variations in case definitions may result in over- or underdiagnosis relative to our findings. Additionally, the transition to usage of *International Classification of Diseases, 10th Revision* may affect the accuracy of coding and merits additional evaluation once sufficient data exist with these codes. The performance of the hospitals in this sample did not associate with measured organizational factors, but other factors not measured in this analysis may affect coding performance, such as payer mix, coding policies, and training. Further evaluation will be needed to identify factors associated with high-accuracy coding.

## CONCLUSIONS

Several important perinatal clinical risk factors and outcomes can reliably

be identified in an administrative claims database, which may allow for the expansion of health services research investigations using these more readily available data. The successful use of administrative diagnosis and procedure codes needs to be based on a good understanding of what each code intends to capture and the consequences for the specific research question to be answered. Caution is needed when using administrative claims data for subjective or difficult-to-define diagnoses and when using infant records to identify maternal conditions. With these findings, we also highlight the opportunity for data quality improvement efforts, because the ability for accurate, comprehensive, and timely extraction of administrative inpatient data will be key to their usefulness as quality metrics.

## ACKNOWLEDGMENT

We acknowledge Beate Danielsen, PhD, for lending her expertise in data linkage.

## ABBREVIATIONS

CPQCC:	California Perinatal Quality Care Collaborative
ICD-9-CM:	<i>International Classification of Diseases, Ninth Revision, Clinical Modification</i>
ICD-9-PCS:	<i>International Classification of Diseases, Ninth Revision, Procedure Coding System</i>
MCC:	Matthews correlation coefficient
NPV:	negative predictive value
OSHPD:	Office of Statewide Health Planning and Development
PDA:	patent ductus arteriosus
PPV:	positive predictive value
ROP:	retinopathy of prematurity



**FINANCIAL DISCLOSURE:** The authors have indicated they have no financial relationships relevant to this article to disclose.

**FUNDING:** Supported by the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (R01 HD083368 [PI: Profit] and R01 HD084667 [PI: Profit]) and the Stanford Child Health Research Institute. Funded by the National Institutes of Health (NIH).

**POTENTIAL CONFLICT OF INTEREST:** The authors have indicated they have no potential conflicts of interest to disclose.

**COMPANION PAPER:** A companion to this article can be found online at [www.pediatrics.org/cgi/doi/10.1542/peds.2018-3293](http://www.pediatrics.org/cgi/doi/10.1542/peds.2018-3293).

## REFERENCES

1. Benchimol EI, Smeeth L, Guttman A, et al; RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med*. 2015;12(10):e1001885
2. Grosse SD, Boulet SL, Amendah DD, Oyeku SO. Administrative data sets and health services research on hemoglobinopathies: a review of the literature. *Am J Prev Med*. 2010;38(suppl 4):S557–S567
3. Romano PS, Chan BK, Schembri ME, Rainwater JA. Can administrative data be used to compare postoperative complication rates across hospitals? *Med Care*. 2002;40(10):856–867
4. Romano PS, Schembri ME, Rainwater JA. Can administrative data be used to ascertain clinically significant postoperative complications? *Am J Med Qual*. 2002;17(4):145–154
5. Bohensky MA, Jolley D, Pilcher DV, Sundararajan V, Evans S, Brand CA. Prognostic models based on administrative data alone inadequately predict the survival outcomes for critically ill patients at 180 days post-hospital discharge. *J Crit Care*. 2012;27(4):422.e11–422.e21
6. Fishman PA, Hornbrook MC, Meenan RT, Goodman MJ. Opportunities and challenges for measuring cost, quality, and clinical effectiveness in health care. *Med Care Res Rev*. 2004;61(suppl 3):124S–143S
7. Gould JB. The role of regional collaboratives: the California Perinatal Quality Care Collaborative model. *Clin Perinatol*. 2010;37(1):71–86
8. Matthews BW. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim Biophys Acta*. 1975;405(2):442–451
9. Boughorbel S, Jarray F, El-Anbari M. Optimal classifier for imbalanced data using Matthews correlation coefficient metric. *PLoS One*. 2017;12(6):e0177678
10. Ford JB, Roberts CL, Algert CS, Bowen JR, Bajuk B, Henderson-Smart DJ; NICUS Group. Using hospital discharge data for determining neonatal morbidity and mortality: a validation study. *BMC Health Serv Res*. 2007;7:188
11. Weingart SN, Iezzoni LI, Davis RB, et al. Use of administrative data to find substandard care: validation of the complications screening program. *Med Care*. 2000;38(8):796–806
12. Iezzoni LI. Assessing quality using administrative data. *Ann Intern Med*. 1997;127(8, pt 2):666–674
13. Scott I, Youlden D, Coory M. Are diagnosis specific outcome indicators based on administrative data useful in assessing quality of hospital care? *Qual Saf Health Care*. 2004;13(1):32–39
14. Buchmueller TC, Allen ME, Wright W. Assessing the validity of insurance coverage data in hospital discharge records: California OSHPD data. *Health Serv Res*. 2003;38(5):1359–1372
15. Howell EA, Janevic T, Hebert PL, Egorova NN, Balbierz A, Zeitlin J. Differences in morbidity and mortality rates in black, white, and Hispanic very preterm infants among New York City hospitals. *JAMA Pediatr*. 2018;172(3):269–277
16. Henderson ZT, Ernst K, Simpson KR, et al. The National Network of State Perinatal Quality Collaboratives: a growing movement to improve maternal and infant health. *J Womens Health (Larchmt)*. 2018;27(3):221–226
17. Engaging clinicians in improving data quality in the NHS. *New J (Inst Health Rec Inf Manag)*. 2006;47(5–6):32–33
18. Zingmond DS, Ye Z, Ettner SL, Liu H. Linking hospital discharge and death records—accuracy and sources of bias. *J Clin Epidemiol*. 2004;57(1):21–29
19. California Perinatal Quality Care Collaborative. Manual of definitions for infants born in 2017. 2017. Available at: [https://www.cpqcc.org/sites/default/files/FORMS/2017/2017\\_CPQCC\\_Manual\\_of\\_Definitions\\_FINAL\\_01.31.17.pdf](https://www.cpqcc.org/sites/default/files/FORMS/2017/2017_CPQCC_Manual_of_Definitions_FINAL_01.31.17.pdf). Accessed April 13, 2018
20. Vermont Oxford Network. Manual of operations: part 2. Data definitions and infant data forms. 2017. Available at: [https://public.vtoxford.org/wp-content/uploads/2016/08/Manual\\_of\\_Operations\\_Part2\\_v21.pdf](https://public.vtoxford.org/wp-content/uploads/2016/08/Manual_of_Operations_Part2_v21.pdf). Accessed April 23, 2018



# Supplemental Information

**SUPPLEMENTAL TABLE 3** Diagnosis and Procedure Codes Used for Study Definitions

	ICD-9-CM and ICD-9-PCS Codes (Based on Infant Record Unless Mentioned Otherwise)
<b>Risk factors</b>	
Very low birth wt	764.01–764.05, 764.11–764.15, 764.21–764.25, 764.91–764.95, 765.01–765.05, 765.11–765.15
Cesarean delivery	Maternal record: 669.7x, 74.0, 74.1, 74.2, 74.4, 74.5, 74.9x
Multiple delivery	Maternal record: 651.x, V27.2, V27.3, V27.4, V27.5, V27.6, V27.7
Maternal chorioamnionitis <sup>a</sup>	Maternal record: 658.4x, 659.21, 659.31; infant record: 762.7
Maternal hypertension	Maternal record: 642.x; infant record: 760.0
Maternal diabetes	Maternal record: 648.0, 648.8, 790.2, 250.x, 249.x; infant record: 775.0
Fetal distress	Maternal record: 655.7x, 656.3x, 659.7x; infant record: 763.0
Meconium aspiration	770.11, 770.12
PDA	747.0, 38.85
<b>Outcomes</b>	
PDA surgery <sup>b</sup>	38.85, 39.79
Mechanical ventilation	V46.1x, 997.31, V44.0, V55.0, 518.8x, 519.0x, 96.7x, 31.1, 31.2x, 31.74
Pneumothorax	512.0, 512.1, 512.8
Respiratory distress syndrome	769
Chronic lung disease <sup>c</sup>	V46.1x, V44.0, V55.0, 518.83, 518.84, 519.0x, 770.7, 31.2x, 31.74
Necrotizing enterocolitis	557.0, 569.83, 578.x, 777.5x, 777.6
Necrotizing enterocolitis surgery	569.6x, V44.2, V44.3, V45.72, V45.73, V45.75, V55.2, V55.3, 17.3x, 45.6x, 45.7x, 45.8x, 45.9x, 46.1x, 46.2x, 46.3x, 46.7x, 46.9x, 47.x, 50.12, 50.3, 50.4, 50.5x, 51.13, 51.2x, 51.3x, 52.12, 52.5x, 52.6, 52.7, 54.1x, 54.21, 54.5x, 54.7x, 55.24
Grade 3 intraventricular hemorrhage <sup>d</sup>	772.13
Grade 4 intraventricular hemorrhage <sup>d</sup>	772.14
Hypoxic-ischemic encephalopathy <sup>e</sup>	768.7x
Stage 3 ROP <sup>f</sup>	362.25
Stage 4 ROP <sup>f</sup>	362.26
Stage 5 ROP <sup>f</sup>	362.27
ROP surgery <sup>f</sup>	14.2x, 14.5x, 14.9
Extracorporeal life support	39.65, 39.66
In-hospital death	Captured from disposition code in discharge record

<sup>a</sup> Collected in CPQCC starting from 2008.

<sup>b</sup> For infants diagnosed with PDA.

<sup>c</sup> Defined as supplemental oxygen at 36 wk corrected gestational age.

<sup>d</sup> For infants with cranial image by day of life 28.

<sup>e</sup> For infants born at 36 wk completed gestation or later.

<sup>f</sup> For infants with very low birth wt with eye examination.

**SUPPLEMENTAL TABLE 4** Organizational Factors Associated With High-Performing and Low-Performing NICUs

	High-Performing NICUs, Median (IQR) or <i>n</i> (%)	Low-Performing NICUs, Median (IQR) or <i>n</i> (%)	<i>P</i>
Licensed beds	15.5 (7–21)	15.5 (6–22)	>.99
Yearly inborn CPQCC infants	40.6 (22.0–96.6)	68.9 (47.2–98.0)	.63
Boys, %	58.3 (54.4–59.6)	55.4 (53.1–57.6)	.22
Birth wt, g	2052 (1870–2250)	2222 (2069–2383)	.39
Gestational age, wk	32.8 (31.9–33.5)	33.4 (32.8–34.3)	.44
High acuity, %	98.8 (88.0–100)	89.9 (71.6–100)	.31
Major congenital anomaly, %	16.8 (15.4–25.7)	11.1 (7.6–16.0)	.06
Length of stay, d	32.5 (25–37)	24 (15–35)	.26
Mothers with prenatal care, %	98.0 (96.4–98.8)	95.7 (95.0–98.2)	.20
Mothers' identification, %			
Hispanic	47.3 (40.7–76.6)	53.9 (34.3–72.6)	.91
Non-Hispanic white	25.4 (11.5–43.2)	22.5 (12.7–31.2)	.53
Asian American	8.4 (6.5–11.6)	9.1 (2.6–13.4)	.91
African American	6.5 (3.8–7.6)	11.1 (2.6–26.5)	.32
Hospital ownership			.47
Public	0 (0)	2 (20)	
Private	10 (100)	8 (80)	
AAP level			.80
Level 2	4 (40)	4 (40)	
Level 3	1 (10)	2 (20)	
Level 4	5 (50)	4 (40)	

High-performing and low-performing NICUs are defined as the 10 NICUs with the highest and lowest mean MCCs, respectively, when comparing perinatal risk factor and outcome coding between an administrative and clinical database. AAP, American Academy of Pediatrics; IQR, interquartile range.