ARTICLE

The Impact of Statistical Choices on Neonatal Intensive Care Unit Quality Ratings Based on Nosocomial Infection Rates

Henry C. Lee, MD, MS; Alyna T. Chien, MD, MS; Naomi S. Bardach, MD; Ted Clay, MS; Jeffrey B. Gould, MD, MPH; R. Adams Dudley, MD, MBA

Objective: To examine the extent to which performance assessment methods affect the percentage of neonatal intensive care units (NICUs) and very low-birth-weight (VLBW) infants included in performance assessments, the distribution of NICU performance ratings, and the level of agreement in those ratings.

Design: Cross-sectional study based on risk-adjusted nosocomial infection rates.

Setting: NICUs belonging to the California Perinatal Quality Care Collaborative 2007-2008.

Participants: One hundred twenty-six California NICUs and 10 487 VLBW infants.

Main Exposures: Three performance assessment choices: (1) excluding "low-volume" NICUs (those caring for <30 VLBW infants per year) vs a criterion based on confidence intervals, (2) using Bayesian vs frequentist hierarchical models, and (3) pooling data across 1 vs 2 years.

Main Outcome Measures: Proportion of NICUs and patients included in quality assessment, distribution of ratings for NICUs, and agreement between methods using the κ statistic.

Results: Depending on the methods applied, 51% to 85% of NICUs and 72% to 96% of VLBW infants were included in performance assessments, 76% to 87% of NICUs were considered "average," and the level of agreement between NICU ratings ranged from 0.23 to 0.89.

Conclusions: The percentage of NICUs included in performance assessments and their ratings can shift dramatically depending on performance measurement method. Physicians, payers, and policymakers should continue to closely examine which existing performance assessment methods are most appropriate for evaluating pediatric care quality.

Arch Pediatr Adolesc Med. 2011;165(5):429-434

Author Affiliations: Divisions of Neonatology (Dr Lee), General Pediatrics (Dr Bardach), and Pulmonary and Critical Care (Dr Dudley) and Philip R. Lee Institute for Health Policy Studies (Drs Lee, Bardach, and Dudley and Mr Clay), University of California at San Francisco, San Francisco; Division of General Pediatrics, Children's Hospital Boston and Harvard Medical School, Boston, Massachusetts (Dr Chien); Division of Neonatal and Developmental Medicine, Perinatal Epidemiology and Health Outcomes Research Unit, Stanford University, Stanford, California (Dr Gould); and California Perinatal Quality Care Collaborative, Stanford (Drs Lee and Gould).

In the United States, recommended pediatric services are provided less than 50% of the time.[1-3] As a result, millions of children are at risk for adverse events during hospitalizations. The care of very low-birth-weight (VLBW) infants in neonatal intensive care units (NICUs) is no exception. Nosocomial infections are a common adverse event, leading to suboptimal outcomes and higher costs.[4-7] In a group of US NICUs, 11.6% of VLBW infants developed NICU-acquired bloodstream infections.[8] In 2007, 1.5% of births were VLBW, translating to approximately 65 000 NICU-acquired infections annually.[9] NICUs have used quality improvement techniques to address this preventable complication.[10-12]

Efforts to use performance incentives, such as pay-for-performance programs, to promote quality are proliferating. Fifty percent to 80% of state Medicaid programs have pay-for-performance programs,[13-15] and recently 3 states passed laws requiring hospitals to publicly report the quality of inpatient pediatric care.[16-18] However, those implementing these programs may not always seek input from child health experts, may lack pediatric quality metrics, and may not always consider the impact of such programs.[19] Recently, the Children's Health Insurance Program Reauthorization Act provided an unprecedented level of federal investment in the development of pediatric quality measures,[20,21] and the Patient Protection and Affordable Care Act is paving the way for using quality measures in a variety of innovative payment arrangements.[22]

In this context, it is crucial that physicians, policymakers, and health insurers (private and governmental) understand the challenges that face pediatric perfor-

©2011 American Medical Association. All rights reserved.
Downloaded From: http://jamanetwork.com/ on 06/19/2013

mance assessment.[23] Our ability to identify outliers is generally impeded by small volumes and low event rates in pediatrics.[24] It is also compounded by the fact that a large percentage of American children, including neonates, receive services in low-volume settings[25-29]; therefore, excluding low-volume providers from performance assessment leaves the quality of care provided to large numbers of children unmeasured and, potentially, "excuses" poor performers.

In this study, we examined how performance assessment methods affect NICU quality ratings using nosocomial infection rates.[4,7,8,30] The NICU setting may be particularly illustrative of pediatrics performance assessment issues because care of VLBW infants has been deregionalizing, and a significant percentage of these infants may receive care in low-volume centers.[25-29,31-34]

Specifically, we evaluated 3 quality assessment choices. The first choice is whether to exclude "low-volume" providers, those caring for fewer than 30 eligible patients, a convention used by The Joint Commission and Medicare,[35,36] vs a rule based on confidence intervals (CIs). The second choice is whether to use Bayesian vs frequentist estimation. We used hierarchical modeling for either approach, drawing on the performance of all NICUs to estimate individual NICU performance. We did not examine the nonhierarchical frequentist method in which each hospital is evaluated separately, an approach whose weaknesses when dealing with low volume or infrequent observations have been described.[37,38] Hierarchical models are increasingly considered in health care performance measurement, such as for coronary arterial bypass surgery evaluation and by Medicare.[37-40] The third choice involves the measurement of period length. Most quality measurements are assessed annually, but pooling data over a longer period may yield more reliable assessments.

Thus, the first objective of this study was to examine how the previously mentioned methodological choices affect the proportion of NICUs (and corresponding VLBW infants) that are included in performance assessments. The second objective was to show to what extent any given method identifies outliers. The third objective was to describe how well NICU ratings agree with one another across these methodological choices.

METHODS

STUDY DESIGN, DATA SOURCE, AND STUDY PERIOD

The University of California at San Francisco Committee on Human Research approved this study. We used a cross-sectional study design using data from the California Perinatal Quality Care Collaborative (CPQCC),[41] a voluntary collective of 128 NICUs that report demographic and clinical data into a central repository. More than 90% of NICUs in California participated in this collaborative between January 1, 2007, and December 31, 2008. Member NICUs collect data on their patients in a prospective manner identical to that submitted to the Vermont Oxford Network.[41-43] Each record undergoes a variety of range, logic, and missing data checks.

STUDY POPULATION

We included VLBW infants (those weighing 400-1499 g) cared for at member NICUs during their initial hospital course. We excluded infants transferred from another NICU after the first day of life (so that receiving NICUs would not be penalized by infections contracted at transferring institutions).

DEFINITION AND NOSOCOMIAL INFECTION MEASURE SPECIFICATIONS

We defined a nosocomial infection event as sepsis or meningitis based on a positive bacterial or fungal culture obtained after the third day of life, following CPQCC and Vermont Oxford Network procedures.[42,44] Events involving coagulase-negative *Staphylococcus* were included if the infant demonstrated other signs of generalized infection and was given antibiotics for 5 days or more. We risk-adjusted NICU-specific nosocomial infection rates similar to the standard CPQCC protocol using gestational age, Apgar score, sex, small for gestational age, singleton or multiple gestation, congenital malformation, prenatal care, any surgery, and birth location.

METHOD COMBINATIONS

We examined 3 performance assessment method combinations and 2 different data pooling periods. In method combination 1, "excluded and hierarchical frequentist," all NICUs used the hierarchical frequentist statistical approach. NICUs with patient volumes less than 30 were excluded. In method combination 2, "included and hierarchical frequentist," we included all NICUs in model estimation, used the hierarchical frequentist approach, and excluded NICUs whose 95% CI contained the 10th and 90th percentiles of the risk-adjusted rates. In method combination 3, "included and hierarchical Bayesian," we included all NICUs, used the hierarchical Bayesian approach, and excluded NICUs whose 95% CI contained the 10th and 90th percentile of risk-adjusted rates. For each combination, we calculated risk-adjusted nosocomial infection rates and corresponding NICU performance ratings using data that pooled all patients in a single year and a 2-year period. Percentiles of risk-adjusted rates were calculated using NICUs with a volume of 30 or more.

OUTCOMES

The main outcomes of interest were (1) the percentage of NICUs (and resulting proportion of VLBW infants) included in performance assessment, (2) the distribution of NICU performance ratings across 3 levels ("above average," "average," and "below average"), and (3) the agreement in NICU performance ratings across the 3 performance assessment combinations and 2 measurement periods.

ANALYSIS

For the first outcome, we calculated the percentage of NICUs that would be included in performance assessment by each of the 3 main combinations of statistical methods and 2 measurement periods and the percentage of VLBW infants seen in those NICUs.

For the second outcome, we used the 95% CIs for each NICU's nosocomial infection rate compared with the mean for the whole group to determine performance rating. If the risk-adjusted upper and lower values of the 95% CI for a NICU were both higher than the mean, the NICU was considered above

©2011 American Medical Association. All rights reserved.
Downloaded From: http://jamanetwork.com/ on 06/19/2013

average; if both values were lower than the mean, the NICU was considered below average.

CI CALCULATION METHODS

In both methods used for calculating CIs for risk-adjusted rates, a hierarchical logistic model was estimated assuming that the logistic transformation of a patient's risk of infection is estimated by the sum of individual risk factors and a hospital effect, assuming that the set of all hospital effects has a normal distribution. Both methods produced a CI for a pseudo–observed to expected ratio. The CIs for a hospital's observed to expected ratio were multiplied by the overall rate to produce the CI for risk-adjusted rates.

HIERARCHICAL FREQUENTIST

We used Proc GLIMMIX in SAS (version 9.2; SAS Institute Inc, Cary, North Carolina) to perform a deterministic calculation producing an estimated value of each hospital effect (centering around zero) with an SE. These were used to obtain a symmetrical 95% CI in the logistic domain using the following formula: [(estimate ± 1.96) × SE]. The lower confidence limit for the observed to expected ratio was calculated as the ratio of the number of infections predicted with vs without the lower confidence limit of the hospital effect and conversely for the upper limit.

HIERARCHICAL BAYESIAN

We obtained CIs from the same model using a Bayesian analysis software package (WinBUGS version 1.4; BUGS Project, MRC Biostatistics Unit, Institute of Public Health, Cambridge, United Kingdom). This method used Monte Carlo simulation to obtain random estimates of hospital observed to expected ratios directly, calculated as the ratio of events predicted with and without the hospital effect. The prior distribution assumed for the hospital effects was a normal distribution with a mean of zero. In addition, noninformative hyperprior distributions were used as part of the estimation process, which had no effect on the model. The 95% CI was obtained as the 2.5% and 97.5% percentiles. In the Monte Carlo method, estimates vary slightly depending on the arbitrary choice of a random number seed, an effect that becomes smaller as the number of iterations increases. Two consecutive 30 000-iteration runs showed no change in performance group assignments.

AGREEMENT

For the last outcome, we tested agreement in ratings across the performance assessment combinations and measurement periods using a simple unweighted κ, which describes agreement between 2 observations, taking agreement by chance into account. We considered ratings to be in agreement only if there was an exact match (ie, both average, both above average, or both below average). If a NICU was excluded by one method and was included by the other method, that was considered nonagreement. We considered that κ ≥ 0.80 would indicate a high level of agreement.[45-47]

RESULTS

Between January 1, 2007, and December 31, 2008, 126 NICUs participating in the CPQCC admitted 10 732 VLBW eligible infants. The mean gestational age of these infants was 28.0 weeks, and their mean birth weight was 1041 g. The mean nosocomial infection rate was 14.4% (interquartile range, 8.7%-17.9%). **Table 1** displays the distribution of NICUs regarding patient volume and service level based on level of neonatal care.[48]

Table 1. Characteristics of 126 Neonatal Intensive Care Units

Variable	Neonatal Intensive Care Units, No. (%)
Level of care	
Regional	23 (18)
Community	69 (55)
Intermediate	23 (18)
Nonclassified	11 (9)
Average annual patient volume	
<30	60 (48)
30-49	31 (25)
≥50	35 (28)

The 123 NICUs that were in the CPQCC for both years of the analysis and that had complete records of risk-adjustment variables included 10 487 patients. The 3 performance assessment combinations and 2 measurement periods being tested yielded a range of 51% to 85% of NICUs and 72% to 96% of patients included in performance ratings (**Table 2**). Approximately half of the NICUs (55%) would be included in performance assessments if low-volume NICUs were excluded and the measurement period was a single year; using the single-year period, approximately the same number of NICUs (54%) would have their performance assessed using the hierarchical Bayesian approach. The number of NICUs assessed increased by using a 2-year period, with 90% to 96% of infants included in the 2-year measurement periods compared with 72% to 84% of infants for 1-year combinations.

Whether NICUs were considered average, above average, or below average differed depending on performance assessment combination and measurement period. Using 1 year of data, 86% to 87% of NICUs were considered average, 3% above average, and 10% to 11% below average. Using 2 years of data, 76% to 78% of NICUs were considered average, 5% to 7% above average, and 16% to 18% below average.

NICUs included by 1 method and excluded by another were always rated as average when included. This was likely due to a characteristic of hierarchical models in which as the number of patients or infections declines, the estimated CI is shifted toward the characteristics of the overall population of hospitals.

The κ statistic comparing NICU performance ratings between performance assessment methods ranged from 0.23 to 0.89 (**Table 3**). The least amount of agreement came from comparing measurement periods that were a single year with those that were 2 years (κ = 0.23-0.42). The level of agreement was higher (κ = 0.56-0.89) when measurement periods were the same; the highest level of agreement was in the 1-year measurement period between the hierarchical frequentist (method 2) and hierarchical Bayesian (method 3) approaches (κ = 0.89).

©2011 American Medical Association. All rights reserved.
Downloaded From: http://jamanetwork.com/ on 06/19/2013

Table 2. Performance Assessment Inclusion Rates and Performance Assessment Distribution According to Analysis Method[a]

Variable	1 y of Data			2 y of Data		
	Method 1	Method 2	Method 3	Method 1	Method 2	Method 3
NICUs included, No. (%)[b]	68 (55)	63 (51)	67 (54)	100 (81)	90 (73)	104 (85)
Patients included, No. (%)[c]	4333 (84)	3724 (72)	3929 (76)	10 110 (96)	9459 (90)	10 013 (95)
NICU NI performance rating distribution, %						
Above average	3	3	3	6	7	5
Average	87	86	87	78	76	77
Below average	10	11	10	16	18	18

Abbreviations: NI, nosocomial infection; NICU, neonatal intensive care unit.
[a] Combined method 1 is "excluded and hierarchical frequentist"; combined method 2, "included and hierarchical frequentist"; and combined method 3, "included and hierarchical Bayesian."
[b] Total number of eligible NICUs = 123.
[c] Total number of eligible patients for 1 year of data was 5173; 2 years of data, 10 487.

Table 3. Agreement of Performance Ratings (κ) Across Performance Assessment Combinations and Data Pooling Periods[a]

Method	1 y of Data			2 y of Data		
	Method 1	Method 2	Method 3	Method 1	Method 2	Method 3
1 y of data						
Method 1	X	0.56	0.64	0.42	0.37	0.31
Method 2		X	0.89	0.26	0.33	0.23
Method 3			X	0.32	0.37	0.27
2 y of data						
Method 1				X	0.69	0.70
Method 2					X	0.72
Method 3						X

[a] Combined method 1 is "excluded and hierarchical frequentist"; combined method 2, "included and hierarchical frequentist"; and combined method 3, "included and hierarchical Bayesian."

COMMENT

We found that performance assessment methods can have a large effect on the percentage of NICUs and VLBW infants included in quality assessments and on performance ratings. In this sector of the health care system where low-volume providers are relatively common, the proportion of NICUs included in performance assessment and the distribution of ratings shifted depending on the method. Choice of method also affected how many NICUs would be considered average. The ability to differentiate providers is important so that the techniques and strategies being used by high performers can be replicated and the practices of low performers can be understood and bolstered. Agreement between performance ratings ranged from 0.23 to 0.89. This finding underscores the variability in performance assessment methods but also gives insight into the level of consistency one should expect if deciding to shift from one method to another.

The choice of method affects how NICUs are labeled in terms of their performance. We used hierarchical statistical modeling, which tends to label low-volume NICUs as average. In contrast, traditional nonhierarchical frequentist methods tend to label these NICUs as low or high performers. Although the methods examined in this study tend to eliminate erratic changes in a NICU's apparent performance over time, there may be a potential cost of labeling a NICU as average simply because it has persistently low patient volume.

There are 2 main limitations to this study: we studied care in California NICUs, a small segment of the pediatric health care system, and we based performance on a single measure. Although this study focused on NICUs in California, the patient population is subject to a dynamic documented for other NICUs[25,27,31-34] and for hospitals that admit children[24,49]: whereas relatively few hospitals care for large volumes of neonates or children, many hospitals care for relatively small volumes of these patients. Indeed, the limitation of small numbers has been documented in other pediatric measures, including an Agency for Healthcare Research and Quality measure of pediatric nosocomial infection rates.[16,24,50,51] Although we based NICU ranks on a single measure of quality, that quality measure is considered clinically valid and reliable and one of the handful of pediatric quality measures currently available.[52,53]

Being able to assess and compare the quality of health care providers is a cornerstone of quality improvement, pay-for-performance, and public reporting programs. Choosing the appropriate performance assessment method depends not only on the availability of evidence-based quality measures but also on an understanding of the ecology of medical care in subsectors of the health care sys-

©2011 American Medical Association. All rights reserved.
Downloaded From: http://jamanetwork.com/ on 06/19/2013

tem. This study illustrates that the rationale for setting minimum patient volumes in adult hospital comparisons may not fully translate to pediatrics because the proportion of providers excluded may be high. This approach can exclude nearly half of existing NICUs from performance assessments. This could be considered negligent in situations in which it is known that lower volume is significantly associated with lower quality.[26,28,29]

Pooling data over a longer period may be a solution to the problem posed by small volumes. It not only allows the inclusion of more providers but also makes it easier to identify outliers and is less susceptible to differences in statistical approaches. Although this strategy may be critiqued for limiting the capacity to track changes in a timely manner, it can be argued that major quality improvement interventions may take at least 1 year to implement,[54] and 2-year periods can be assessed in a rolling manner such that performance assessments always include the most recent year's data.

We used a definition of nosocomial infection aligned with that of the Vermont Oxford Network. However, there are other relevant definitions. For example, the Centers for Disease Control and Prevention National Healthcare Safety Network specifies infections associated with a central line, using device patient-days as the denominator.[55] Because the requirement of central line placement would reduce the denominator, the effect on exclusion of smaller NICUs may be greater with this approach. The nosocomial infection measure, which is one of the Pediatric Quality Indicators from the Agency for Healthcare Research and Quality, is an attempt to closely approximate the CPQCC definition of nosocomial infection using administrative data.[56] The Joint Commission Perinatal Core Measure set includes health care–associated bloodstream infections in newborns and also relies on administrative data.[53] Administrative data may not accurately identify health care–associated infections as well as prospectively collected clinical data.[57] Further study of alternative definitions of nosocomial infection may underscore the effect on low-volume settings. We found a slight difference in performance assessment inclusion and ratings when comparing hierarchical frequentist and Bayesian analyses (Table 2). This analysis does not use a criterion standard and, therefore, gives no insight into whether a Bayesian vs frequentist approach is preferable. However, considering that the agreement between these methods was relatively high ($\kappa = 0.72$ for 2 years and $\kappa = 0.89$ for 1 year), this choice may not be crucial if hierarchical modeling is used.

Children in the United States deserve the safest and most effective health care that modern medicine has to offer. Recent federal legislation is likely to be a positive force in establishing programs aimed at measuring and promoting pediatric health care quality. Quality ratings of pediatric providers will likely be at the core of these efforts, and it is essential that interested parties understand how even basic performance assessment conventions can affect whether providers are included in quality assessments and performance ratings. In conclusion, different strategies for performance assessment lead to differing inclusion and ratings for NICU comparisons, particularly when low-volume providers are commonplace. Physicians, payers, and policymakers should continue to closely examine the extent to which performance assessment methods affect which pediatric providers are assessed for quality and the ratings they may receive.

Accepted for Publication: January 19, 2011.
Correspondence: Henry C. Lee, MD, MS, Department of Pediatrics, Division of Neonatology, University of California at San Francisco, 533 Parnassus Ave, Room U503, San Francisco, CA 94143 (LeeHC@peds.ucsf.edu).
Author Contributions: Dr Lee had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. *Study concept and design:* Lee, Chien, Gould, and Dudley. *Acquisition of data:* Lee, Chien, Gould, and Dudley. *Analysis and interpretation of data:* Lee, Chien, Bardach, Clay, and Dudley. *Drafting of the manuscript:* Lee, Chien, and Dudley. *Critical revision of the manuscript for important intellectual content:* Lee, Chien, Bardach, Clay, Gould, and Dudley. *Statistical analysis:* Lee, Chien, Clay, and Dudley. *Obtained funding:* Lee and Dudley. *Administrative, technical, and material support:* Gould and Dudley. *Study supervision:* Dudley.
Financial Disclosure: None reported.
Funding/Support: This project was supported by NIH/NCRR/OD UCSF-CTSI grant KL2 RR024130 (Dr Lee), by an Investigator Award in Health Policy from the Robert Wood Johnson Foundation (Dr Dudley), and by the California Hospital Assessment and Reporting Taskforce (Dr Dudley). Data were provided by the CPQCC.
Disclaimer: The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

REFERENCES

1. Schuster MA, McGlynn EA, Brook RH. How good is the quality of health care in the United States? 1998. *Milbank Q.* 2005;83(4):843-895.
2. Schuster MA, McGlynn EA. *Measuring the Quality of Care in Pediatrics.* Philadelphia, PA: Lippincott Williams & Wilkins; 1999.
3. Mangione-Smith R, DeCristofaro AH, Setodji CM, et al. The quality of ambulatory care delivered to children in the United States. *N Engl J Med.* 2007;357(15):1515-1523.
4. Brady MT. Health care-associated infections in the neonatal intensive care unit. *Am J Infect Control.* 2005;33(5):268-275.
5. Stoll BJ, Hansen NI, Adams-Chapman I, et al; National Institute of Child Health and Human Development Neonatal Research Network. Neurodevelopmental and growth impairment among extremely low-birth-weight infants with neonatal infection. *JAMA.* 2004;292(19):2357-2365.
6. Payne NR, Carpenter JH, Badger GJ, Horbar JD, Rogowski J. Marginal increase in cost and excess length of stay associated with nosocomial bloodstream infections in surviving very low birth weight infants. *Pediatrics.* 2004;114(2):348-355.
7. Fanaroff AA, Korones SB, Wright LL, et al; The National Institute of Child Health and Human Development Neonatal Research Network. Incidence, presenting features, risk factors and significance of late onset septicemia in very low birth weight infants. *Pediatr Infect Dis J.* 1998;17(7):593-598.
8. Sohn AH, Garrett DO, Sinkowitz-Cochran RL, et al; Pediatric Prevention Network. Prevalence of nosocomial infections in neonatal intensive care unit patients: results from the first national point-prevalence survey. *J Pediatr.* 2001;139(6):821-827.
9. Hamilton BE, Martin JA, Ventura SJ. Births: preliminary data for 2007. *Natl Vital Stat Rep.* 2009;57(12):1-23.
10. Aly H, Herson V, Duncan A, et al. Is bloodstream infection preventable among premature infants? A tale of two cities. *Pediatrics.* 2005;115(6):1513-1518.
11. Lee SK, Aziz K, Singhal N, et al. Improving the quality of care for infants: a cluster randomized controlled trial. *CMAJ.* 2009;181(8):469-476.

©2011 American Medical Association. All rights reserved.
Downloaded From: http://jamanetwork.com/ on 06/19/2013

12. O'Grady NP, Alexander M, Dellinger EP, et al; Hospital Infection Control Practices Advisory Committee, Center for Disease Control and Prevention. Guidelines for the prevention of intravascular catheter-related infections. *Pediatrics*. 2002;110(5):e51.
13. Kurhmerker K, Hartman T. Pay-for-performance in state Medicaid programs: a survey of state Medicaid directors and programs. http://www.commonwealthfund.org/Content/Publications/Fund-Reports/2007/Apr/Pay-for-Performance-in-State-Medicaid-Programs--A-Survey-of-State-Medicaid-Directors-and-Programs.aspx. Accessed April 23, 2007.
14. Kurhmerker K, Hartman T. Pay-for-performance in state Medicaid programs: appendix B: state pay-for-performance program summaries. The Commonwealth Fund Web site. http://www.commonwealthfund.org/~/usr_doc/Kuhmerker_P4PstateMedicaidprogsappendixB.pdf?section=4039. Accessed April 23, 2007.
15. Chien AT, Li Z, Rosenthal MB. Improving timely childhood immunizations through pay for performance in Medicaid-managed care. *Health Serv Res*. 2010;45 (6, pt 2):1934-1947.
16. Center for Health Statistics. Texas Health Care Information Collection: quality of children's care in Texas hospitals, 2008. http://www.dshs.state.tx.us/THCIC/Publications/Hospitals/PDIReport/PDIReport.shtm. Accessed March 2009.
17. Association of Florida Children's Hospitals. Recommendation from the Statewide Workgroup on Pediatric Data provided on AHCA's consumer Web site. http://ahca.myflorida.com/SCHS/doc/2006_10_19/Tab%20B%20Pediatric%20Recommendations.pdf. Accessed April 2009.
18. Department of Banking, Insurance, Securities and Health Care Administration. Vermont hospital report card 2008 comparison report: volume and mortality for selected procedures. http://www.bishca.state.vt.us/sites/default/files/Volume-Mortality2008.pdf. Accessed February 28, 2011.
19. Chien AT, Colman MW, Ross LF. Qualitative insights into how pediatric pay-for-performance programs are being designed. *Acad Pediatr*. 2009;9(3):185-191.
20. Children's Health Insurance Program Reauthorization Act of 2009. Public Law 111-3. Feb 4, 2009. http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=111_cong_public_laws&docid-f:pub1003.111.pdf. Accessed August 19, 2010.
21. Dougherty D, Simpson LA. Measuring the quality of children's health care: a prerequisite to action. *Pediatrics*. 2004;113(1, pt 2):185-198.
22. Patient Protection and Affordable Care Act 2010. http://www.govtrack.us/congress/billtext.xpd?bill=h111-3590. Accessed August 19, 2010.
23. Chien AT, Dudley RA. Pay-for-performance in pediatrics: proceed with caution. *Pediatrics*. 2007;120(1):186-188.
24. Bardach NS, Chien AT, Dudley RA. Small numbers limit the use of the inpatient pediatric quality indicators for hospital comparison. *Acad Pediatr*. 2010;10 (4):266-273.
25. Howell EM, Richardson D, Ginsburg P, Foot B. Deregionalization of neonatal intensive care in urban areas. *Am J Public Health*. 2002;92(1):119-124.
26. Phibbs CS, Baker LC, Caughey AB, Danielsen B, Schmitt SK, Phibbs RH. Level and volume of neonatal intensive care and mortality in very-low-birth-weight infants. *N Engl J Med*. 2007;356(21):2165-2175.
27. Yeast JD, Poskin M, Stockbauer JW, Shaffer S. Changing patterns in regionalization of perinatal care and the impact on neonatal mortality. *Am J Obstet Gynecol*. 1998;178(1, pt 1):131-135.
28. Bartels DB, Wypij D, Wenzlaff P, Dammann O, Poets CF. Hospital volume and neonatal mortality among very low birth weight infants. *Pediatrics*. 2006;117 (6):2206-2214.
29. Phibbs CS, Bronstein JM, Buxton E, Phibbs RH. The effects of patient volume and level of care at the hospital of birth on neonatal mortality. *JAMA*. 1996; 276(13):1054-1059.
30. Stover BH, Shulman ST, Bratcher DF, Brady MT, Levine GL, Jarvis WR; Pediatric Prevention Network. Nosocomial infection rates in US children's hospitals' neonatal and pediatric intensive care units. *Am J Infect Control*. 2001;29(3): 152-157.
31. Mehta S, Atherton HD, Schoettker PJ, Hornung RW, Perlstein PH, Kotagal UR. Differential markers for regionalization. *J Perinatol*. 2000;20(6):366-372.
32. Powell SL, Holt VL, Hickok DE, Easterling T, Connell FA. Recent changes in delivery site of low-birth-weight infants in Washington: impact on birth weight-specific mortality. *Am J Obstet Gynecol*. 1995;173(5):1585-1592.
33. Haberland CA, Phibbs CS, Baker LC. Effect of opening midlevel neonatal intensive care units on the location of low birth weight births in California. *Pediatrics*. 2006;118(6):e1667-e1679. http://pediatrics.aappublications.org/cgi/content/full/peds.2006-0612v1. Accessed January 17, 2011.
34. Gould JB, Marks AR, Chavez G. Expansion of community-based perinatal care in California. *J Perinatol*. 2002;22(8):630-640.
35. Centers for Medicare & Medicaid Services. Physician quality reporting initiative: establishment of alternative reporting periods and reporting criteria. http://www.cms.hhs.gov/PQRI/Downloads/2008PQRIalterrptperiods.pdf. Accessed January 18, 2011.
36. The Joint Commission. Guidelines for publicizing hospital national quality improvement goals. http://www.jointcommission.org/guidelines_for_publicizing_nqigs. Accessed February 28, 2011.
37. Shahian DM, Normand SL. Low-volume coronary artery bypass surgery: measuring and optimizing performance. *J Thorac Cardiovasc Surg*. 2008;135(6): 1202-1209.
38. Shahian DM, Torchiana DF, Shemin RJ, Rawn JD, Normand SL. Massachusetts cardiac surgery report card: implications of statistical methodology. *Ann Thorac Surg*. 2005;80(6):2106-2113.
39. Austin PC. A comparison of Bayesian methods for profiling hospital performance. *Med Decis Making*. 2002;22(2):163-172.
40. Austin PC, Naylor CD, Tu JV. A comparison of a Bayesian vs. a frequentist method for profiling hospital performance. *J Eval Clin Pract*. 2001;7(1):35-45.
41. Wirtschafter DD, Powers RJ. Organizing regional perinatal quality improvement: global considerations and local implementation. *NeoReviews*. 2004; 5:e50-e59.
42. CPQCC Network Database Manual of Definitions for Infants Born in 2010. Version 1.10. http://cpqcc.org/data/cpqcc_downloads. Accessed February 28, 2011.
43. Horbar JD. The Vermont-Oxford Neonatal Network: integrating research and clinical practice to improve the quality of medical care. *Semin Perinatol*. 1995;19 (2):124-131.
44. Vermont Oxford Network. Vermont Oxford Network Database: manual of operations for infants born in 2009. http://www.vtoxford.org/tools/2009ManualofOperationswithindex13_2.pdf. Accessed September 8, 2010.
45. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174.
46. Altman DG. *Practical Statistics for Medical Research*. London, UK: Chapman & Hall; 1991.
47. Yung M, South M, Byrt T. Evaluation of an asthma severity score. *J Paediatr Child Health*. 1996;32(3):261-264.
48. Wirtschafter DD, Danielsen BH, Main EK, et al; California Perinatal Quality Care Collaborative. Promoting antenatal steroid use for fetal maturation: results from the California Perinatal Quality Care Collaborative. *J Pediatr*. 2006;148(5):606-612.
49. Chamberlain LJ, Chan J, Mahlow P, Huffman LC, Chan K, Wise PH. Variation in specialty care hospitalization for children with chronic conditions in California. *Pediatrics*. 2010;125(6):1190-1199.
50. Florida Agency for Health Care Administration. FloridaHealthFinder.gov Web site. http://www.floridahealthfinder.gov/CompareCare/CompareFacilities.aspx. Accessed January 11, 2011.
51. Vermont's inpatient volume and mortality indicators for selected procedures. Department of Banking, Insurance, Securities & Health Care Administration. http://www.bishca.state.vt.us/health-care/hospitals-health-care-practitioners/2009-volume-mortality. Accessed December 6, 2010.
52. National Quality Forum. *National Voluntary Consensus Standard for Perinatal Care: Peforrmance Measure Specifications*. Washington, DC: National Quality Forum; October 20, 2008. Accessed January 18, 2011.
53. The Joint Commission. Specifications manual for Joint Commission national quality core measures (2010A2). http://manual.jointcommission.org/releases/TJC2010A/.
54. Wagner EH, Glasgow RE, Davis C, et al. Quality improvement in chronic illness care: a collaborative approach. *Jt Comm J Qual Improv*. 2001;27(2):63-80.
55. Centers for Disease Control and Prevention. Central line-associated bloodstream infection (CLABSI) event. http://www.cdc.gov/nhsn/PDFs/pscManual/4PSC_CLABScurrent.pdf. Accessed February 28, 2011.
56. Agency for Healthcare Research and Quality. Measures of pediatric health care quality based on hospital administrative data: the pediatric quality indicators: neonatal indicator appendix. April 17, 2008. http://www.qualityindicators.ahrq.gov/downloads/pdi/pdi_measures_neonatal_appendix.pdf. Accesssed January 11, 2011.
57. Sherman ER, Heydon KH, St John KH, et al. Administrative data fail to accurately identify cases of healthcare-associated infection. *Infect Control Hosp Epidemiol*. 2006;27(4):332-337.

©2011 American Medical Association. All rights reserved.
Downloaded From: http://jamanetwork.com/ on 06/19/2013