ARTICLE

Correlation of Neonatal Intensive Care Unit Performance Across Multiple Measures of Quality of Care

Jochen Profit, MD, MPH; John A. F. Zupancic, MD, ScD; Jeffrey B. Gould, MD, MPH; Kenneth Pietz, PhD; Marc A. Kowalkowski, MS; David Draper, PhD; Sylvia J. Hysong, PhD; Laura A. Petersen, MD, MPH

Objectives: To examine whether high performance on one measure of quality is associated with high performance on others and to develop a data-driven explanatory model of neonatal intensive care unit (NICU) performance.

Design: We conducted a cross-sectional data analysis of a statewide perinatal care database. Risk-adjusted NICU ranks were computed for each of 8 measures of quality selected based on expert input. Correlations across measures were tested using the Pearson correlation coefficient. Exploratory factor analysis was used to determine whether underlying factors were driving the correlations.

Setting: Twenty-two regional NICUs in California.

Patients: In total, 5445 very low-birth-weight infants cared for between January 1, 2004, and December 31, 2007.

Main Outcomes Measures: Pneumothorax, growth velocity, health care–associated infection, antenatal corticosteroid use, hypothermia during the first hour of life, chronic lung disease, mortality in the NICU, and discharge on any human breast milk.

Results: The NICUs varied substantially in their clinical performance across measures of quality. Of 28 unit-level correlations, 6 were significant ($\rho < .05$). Correlations between pairs of measures of quality of care were strong ($\rho \geq .5$) for 1 pair, moderate (range, $\rho \geq .3$ to $\rho < .5$) for 8 pairs, weak (range, $\rho \geq .1$ to $\rho < .3$) for 5 pairs, and negligible ($\rho < .1$) for 14 pairs. Exploratory factor analysis revealed 4 underlying factors of quality in this sample. Pneumothorax, mortality in the NICU, and antenatal corticosteroid use loaded on factor 1; growth velocity and health care–associated infection loaded on factor 2; chronic lung disease loaded on factor 3; and discharge on any human breast milk loaded on factor 4.

Conclusion: In this sample, the ability of individual measures of quality to explain overall quality of neonatal intensive care was modest.

JAMA Pediatr. 2013;167(1):47-54.
Published online November 12, 2012.
doi:10.1001/jamapediatrics.2013.418

QUALITY OF CARE PROVIDED by providers is increasingly scrutinized in an attempt to increase efficiency and improve the quality of patient care.[1,2] In other areas of medicine, performance measurement and financial incentives are common.[3-5] In the neonatal intensive care unit (NICU) setting, multistakeholder health care organizations (such as the National Quality Forum[6]) and payers of health care are promoting performance assessments of perinatal care providers.

Two facets of performance measurement have received little attention. First, is it fair to draw conclusions regarding institutional performance based on a single or limited set of measures of quality of care? Conclusions based on a small or limited set assume that measured aspects of quality reflect unmeasured aspects of care. However, a study[7] of hospital quality assessments based on hospitalwide mortality rates alone found substantial discrepancies in performance based on the methods used to calculate mortality rates. This calls into question whether it is valid to draw conclusions about quality of care based on hospitalwide mortality rates. In the NICU setting, good performance on one measure of quality (eg, the proportion of infants with chronic lung disease) is assumed to indicate good performance on related measures of quality (eg, duration of mechanical ventilation) and on unrelated measures (eg, rates of health care–associated infection).

For editorial comment see page 89

Author Affiliations are listed at the end of this article.

The use of a limited set of measures of quality of care for comparative performance measurement would be supported if NICU performance was strongly correlated across multiple measures of quality of care. However, in other areas of health care, studies[8-12] have found weak or no correlation across measures of quality of care. If intrainstitutional correlations among measures of quality of care are weak and performance is inconsistent, then inferences about quality from 1 or a few measures of quality are likely uninformative and potentially misleading.[13] Instead, quality should be assessed by combining multiple measures of quality into 1 or more composite indicators of quality.[14]

Second, should quality improvement efforts be directed toward individual measures of quality or toward building more tightly connected systems of care so that performance can be based on several measures of quality simultaneously? Traditional approaches to quality improvement have typically addressed individual measures sequentially.[15,16] In many instances, this has promoted better, safer care, but often gains have been temporary. A growing body of literature suggests that sustained and widespread improvements in quality require changes to the system in which care occurs. For example, improvements in unit safety culture, which varies widely across NICUs,[17] have been linked to lasting improvements.[18,19] The system supporting care delivery is interconnected with quality of care provided. Therefore, correlations between measures of quality might be interpreted to reflect the degree of care systems integration. Weak correlations might suggest a low degree of systems integration, in which care processes are largely functionally independent.[9] Such a finding might signal the need for interventions, such as improvements in safety culture[20] or composite measurement of quality,[14,21] that could more broadly affect performance.

Neonatal intensive care presents a natural laboratory to test whether comparative performance measurement should be approached via limited or expanded sets of measures of quality of care. Specifically, high-quality clinical data are being collected by the California Perinatal Quality Care Collaborative (CPQCC) and other quality-of-care consortia. Our group has been working with CPQCC data to develop a composite indicator of neonatal intensive care quality provided to very low-birth-weight infants, the Baby-MONITOR.[21,22] This study uses CPQCC data and 8 measures of quality that have been selected for inclusion in the Baby-MONITOR to examine the consistency of NICU performance rankings. We hypothesized that correlations of NICU rankings across measures of quality would be at least moderate. The specific objectives of this study were to examine whether high performance on one measure of quality is associated with high performance on others and to develop a data-driven explanatory model for overall NICU performance measurement.

METHODS

OVERVIEW

The CPQCC[15] is a multistakeholder group of public and private obstetric and neonatal providers, health care purchasers, public health professionals, and private sector health industry specialists, committed to improving care and outcomes for the state's pregnant mothers and newborns. The collaborative includes more than 130 member hospitals, of which 24 are designated as regional centers. This roster accounts for most of the preterm infants requiring critical care in California.

Table 1. Study Exclusion Criteria and Rationale

Exclusion Criterion	Rationale
Transfer out for reasons other than convalescent or chronic care	Minimize exclusion of patients, maximize information content
Gestational age <25 wk	Minimize systematic bias at the border of viability
Body weight >1500 g, major congenital anomalies[a]	Maximize comparability across NICUs
Death within 12 h of birth, transfer in after age 3 d, readmission or death after transfer out	Ensure performance can be ascribed to NICU under evaluation

Abbreviation: NICU, neonatal intensive care unit.
[a] Associated with an elevated mortality risk (see http://cpqcc.org/data/cpqcc_downloads for a detailed definition in the *Manual of Definitions*).

PATIENT SELECTION

In total, 5445 very low-birth-weight infants cared for at 22 of 24 California level III regional centers between January 1, 2004, and December 31, 2007, met inclusion criteria for the study. Of these centers, 15 are designated as level IIID on the basis of open heart surgery performance, and the remainder are designated as level IIIC.[23] We used multiyear analysis because of the few very low-birth-weight infants cared for in some institutions. Detailed descriptions of measure selection, definition, and exclusion criteria have been published elsewhere[21] and are summarized in **Table 1**. Additional technical details are provided in the eAppendix (http://www.jamapeds.com).

DATA

Dependent Variables

We chose 8 quality-of-care measures that had been selected by an expert panel in a modified Delphi experiment for inclusion in the Baby-MONITOR and that have subsequently been confirmed by a sample of clinical neonatologists. Measure definitions were derived from standard CPQCC and Vermont Oxford Network algorithms.[21,24] Measures included the following: (1) antenatal corticosteroid use, (2) hypothermia (<36°C) during the first hour of life, (3) nonsurgically induced pneumothorax, (4) health care–associated bacterial or fungal infection, (5) survival to discharge or to 36 weeks' gestational age with chronic lung disease (need for oxygen therapy or mechanical ventilation at 36 weeks' gestational age), (6) discharge on any human breast milk, (7) mortality in the NICU during the birth hospitalization, and (8) growth velocity. Growth velocity was determined according to a logarithmic function.[25] We aligned all variables so that a higher value represents a better outcome. Statistical modeling (described herein) for this analysis required transformation of continuous variables into categorical ones. Therefore, we empirically dichotomized growth velocity into high- and low-growth groups based on the median velocity of 12.4 g/kg/d derived from the 95% central sample. The denominators for the variables differ slightly. For example, infants who died in the NICU or who survived but remained in the NICU for more than 6 months are not included in the denominator for the breast milk variable.

Independent Variables

We applied CPQCC standard operational definitions for all independent variables. Patients were grouped into gestational age at birth strata of $25^{0/7}$ to $27^{6/7}$, $28^{0/7}$ to $29^{6/7}$, and $30^{0/7}$ or more weeks based on similar patient numbers between groups. Apgar score was categorized as 3 or less, 4 to 6, or greater than 6.

STATISTICAL ANALYSIS

Basic descriptive analyses examined the variation in unadjusted measures across sites. Hospital-level data included each level III NICU as the unit of analysis. To adjust for confounding due to differences in case mix, we developed risk adjustment models for each measure. For each one, we selected a set of candidate variables based on reported associations in the literature or clinical relevance, and we tested for associations with the outcome of interest in univariate analyses using the Fisher exact test for categorical variables and, based on the underlying variable distribution, the *t* test or the 2-sample Wilcoxon signed rank test. Variables associated at a significance level of $P \leq .25$ were entered into a logistic regression model, and variables associated at a significance level of $P > .05$ were successively removed from the model after checking the log-likelihood ratio test for contribution to model fit.[26]

To rank NICU performance on each measure of quality of care, we used a method that was developed by Draper and Gittoes[27] for use in the United Kingdom educational system and which is relevant and valid in any profiling setting with dichotomous outcomes. For each NICU and for each measure of quality of care, a *z* score was computed as the observed rate minus the expected rate, divided by its estimated standard error. The NICU's expected value was computed as a weighted mean of the rate (eg, the survival rate) in the overall database for all levels of the risk adjustment variables.

Objective 1: Consistency of High Performance

We used 2 approaches to examine the degree to which superior performance on one key measure of quality (survival) was associated with superior performance on the other measures.[9] First, we ranked NICU performance on each measure according to its *z* score and calculated correlations of the *z* scores using the Pearson correlation coefficient. Correlations were rated as weak, moderate, or strong according to conventional thresholds.[28] Second, we compared the distribution of being in the top 4 ranks across measures to a binomial distribution using a χ^2 test. A test result that is statistically nonsignificant indicates that the hypothesis of independence cannot be rejected.

Objective 2: Development of a Model of Overall NICU Performance

We performed an exploratory factor analysis to determine whether underlying factors were driving the correlations. Factor loadings in excess of 0.5 were used to classify variables into factors. For all analyses, $P < .05$ was considered statistically significant. Detailed information on model building, the method by Draper and Gittoes,[27] and factor analysis is given in the eAppendix.

HUMAN STUDY COMPLIANCE

The CPQCC data are collected for quality improvement and meet the criteria for deidentified data. The data set is then further deidentified with respect to hospital for use as a research data set. The study was approved by the CPQCC and by the Baylor College of Medicine institutional review board.

RESULTS

CHARACTERISTICS OF INFANTS AND NICUs

Table 2 gives characteristics of the study sample. The means for the measures of quality of care are adjusted for illness severity at birth.

NICU *z* SCORES AND RANKS ACROSS MEASURES

Table 3 lists *z* scores of performance on each variable (the standardized observed minus expected rate), with the NICUs labeled A through V in descending order of survival. A *z* score of zero indicates that the observed results on the measures of quality of care equal the expected (ie, risk-adjusted) results. A positive number indicates that performance is better than expected. We found substantial variation within measures of quality of care between NICUs, except for pneumothorax. A separate analysis using random-effects models showed significant NICU-level variation for all outcomes except pneumothorax (data are available from the author on request).

OBJECTIVE 1: CONSISTENCY OF HIGH PERFORMANCE

Correlation Among Measures of Quality of Care

Table 4 gives the NICU-level correlation matrix among measures of quality of care. Of 28 unit-level correlations, 6 were significant ($\rho < .05$). Correlations between pairs of measures of quality of care were strong ($\rho \geq .5$) for 1 pair, moderate (range, $\rho \geq .3$ to $\rho < .5$) for 8 pairs, weak (range, $\rho \geq .1$ to $\rho < .3$) for 5 pairs, and negligible ($\rho < .1$) for 14 pairs.

Consistency of High Performance Across Measures of Quality

We found little consistency of high performance between NICUs. The number of times that NICUs were among the top 4 ranks (a high performer) for the 8 measures of quality of care ranged from 0 (never among the top 4 ranks) to 4 (being in the top 4 ranks for 4 of 8 measures). **Figure 1** shows the observed and expected distribution under an assumption that high performance on different measures occurs at random (according to a binomial distribution in which the probability of success on each trial is 4/24 = 0.17 and the 8 trials are independent). The observed distribution from the random binomial distribution was not statistically different ($P > .9$). Nevertheless, the sum of ranks (**Figure 2**) across measures of quality of care suggests that hospitals performing well on survival tend to do well on other measures of quality.

Table 2. Characteristics of Infants and NICUs

Characteristic	Infant Level (n = 5445)	NICU Level (n = 22)
	Mean (SD) [Range]	Annual Mean (SD) [Range], No. of Infants per Hospital
Gestational age at birth, wk	28.5 (2.4) [25-38]	28.6 (0.3) [28-29]
Birth weight, g	1092 (259) [334-1499]	1095 (25.6) [1035-1133]
Apgar score at 5 min	6.6 (1.1) [0.0-7.0] (n = 5406)	6.6 (0.2) [6.2-6.8]
	No. (%)	Annual Mean (SD) [Range], No. of Infants per Hospital
Small for gestational age	1526 (28.0)	28.3 (6.3) [13.4-41.1]
Female sex	2670 (49.0)	49.0 (4.8) [36.3-59.6]
Cesarean delivery	3994 (73.4)	72.3 (7.4) [55.3-86.6]
Multiple gestation	1641 (30.1)	28.5 (8.0) [12.5-44.1]
Any prenatal care	5184 (95.2)	95.5 (4.2) [83.8-100.0]
Infant race/ethnicity		
Non-Hispanic black	647 (11.9)	12.3 (8.3) [2.1-31.3]
Non-Hispanic white	1721 (31.6)	28.4 (14.0) [0.0-46.0]
Hispanic	2343 (43.0)	45.7 (20.4) [19.1-97.9]
Asian	587 (10.8)	10.0 (7.3) [0.0-30.9]
Other	147 (2.7)	3.6 (6.1) [0.0-28.9]
Very low-birth-weight admissions		248 (161) [47-655]
Inborn infants	3838 (70.5)	72.2 (35.5) [0.0-97.6]
Infants transferred out	1257 (23.0) (n = 5456)	22.2 (19.6) [1.6-70.9]
Transferred for convalescent care	1041 (19.1) (n = 5456)	17.7 (19.0) [0.0-69.4]
Death after transfer	5 (0.1) (n = 5456)	0.2 (0.5) [0.0-2.1]
Readmission after transfer	115 (2.1) (n = 5456)	2.6 (2.8) [0.0-10.3]
Baby-MONITOR Measure of Quality of Care[22]	**No. (%) (No. Based on Measure Definition)**	**Risk-Adjusted Mean (SD) [Range], %**
Survival	4955 (94.8) (n = 5229)	94.2 (3.3) [85.7-98.6]
Any antenatal corticosteroid use	4270 (78.4) (n = 5445)	76.5 (12.2) [46.3-91.6]
No hypothermia <36.0°C at 1 h of life	2228 (81.3) (n = 2739)	78.6 (11.5) [58.8-97.9]
No pneumothorax	5244 (96.5) (n = 5437)	96.4 (1.9) [91.3-100.0]
No health care–associated infection	4274 (81.3) (n = 5259)	82.1 (6.4) [73.2-93.4]
High growth velocity >12.4 g/kg/d	2355 (50.6) (n = 4651)	51.4 (14.1) [23.4-83.5]
No chronic lung disease at 36 wk gestational age	3607 (74.6) (n = 4837)	74.6 (9.7) [50.3-89.9]
Discharge on any human breast milk	2310 (59.8) (n = 3864)	61.2 (18.9) [22.2-97.5]

Abbreviations: NA, not applicable; NICU, neonatal intensive care unit.

OBJECTIVE 2: DEVELOPMENT OF A MODEL OF OVERALL NICU PERFORMANCE

Exploratory factor analysis revealed 4 underlying factors of quality in this sample (**Table 5**). Pneumothorax, mortality in the NICU, and antenatal corticosteroid use loaded on factor 1; growth velocity and health care–associated infection loaded on factor 2; chronic lung disease loaded on factor 3; and discharge on any human breast milk loaded on factor 4. Hypothermia during the first hour of life did not load on any factor. These factors might be clinically interpreted as follows: factor 1 may reflect the quality of perinatal care because the consequences of good perinatal care are low rates of pneumothorax and high survival; factor 2 may reflect the quality of supporting healthy development, which would be endangered by poor growth velocity and health care–associated infection; factor 3 may represent the quality of respiratory care because good care results in low rates of chronic lung disease; and factor 4 may reflect maternal involvement, which is key to achieving high rates of discharge on any human breast milk.

COMMENT

In this article, we examined NICU performance on 8 measures of quality of care. Except for the variable measuring pneumothorax, we found significant variation in clinical processes and outcomes between NICUs within and across each measure of quality. Correlations between most measures of quality were modest, and performance on one measure of quality had little predictive accuracy regarding performance on another. The only exception was high growth velocity and the absence of health care–associated infection, which were reasonably correlated. An exploratory factor analysis revealed 4 underlying factors of quality in this sample.

Our results have important implications for the comparative performance measurement endeavor. Given the modest correlations among measures of quality of care and the inconsistency among relative performances, one should not infer overall NICU quality based on a single measure or a few measures of quality. Our findings call into question the assumption that this measurement approach will lead to widespread improvements in quality, a method that underlies current benchmarking ef-

Table 3. Risk-Adjusted *z* Scores by NICU and Measure of Quality of Care, Showing Within-NICU and Between-NICU Variation[a]

NICU	Survival	Any Antenatal Corticosteroid Use	No Hypothermia at 1 h of Life	No Pneumothorax	No Health Care–Associated Infection	High Growth Velocity	No Chronic Lung Disease at 36 wk Gestational Age	Discharge on Any Human Breast Milk	Mean (SD)
A	5.69	–0.20	2.40	1.36	–0.05	0.89	2.24	–3.43	1.11 (2.61)
B	2.66	5.52	–5.92	1.67	–1.91	–2.20	1.86	–0.25	0.18 (3.53)
C	1.26	–0.81	0.40	–0.14	3.39	0.60	–3.93	–0.45	0.04 (2.07)
D	1.20	4.21	–2.60	0.01	2.84	1.95	2.47	4.45	1.82 (2.31)
E	1.15	2.46	–0.45	2.22	0.02	5.18	–3.42	–6.61	0.07 (3.68)
F	0.92	–2.83	–3.46	–0.11	–0.25	1.25	0.08	–7.02	–1.43 (2.81)
G	0.76	1.22	–0.19	–1.78	0.27	0.97	2.94	–1.65	0.32 (1.55)
H	0.45	6.52	4.02	–0.66	–0.88	–2.57	5.21	1.35	1.68 (3.23)
I	0.06	–0.26	3.66	1.54	2.65	0.17	–1.36	5.20	1.46 (2.23)
J	–0.07	–4.50	–4.20	–0.48	0.77	–0.50	1.99	1.94	–0.63 (2.49)
K	–0.10	4.97	–1.83	1.22	–2.40	–2.60	6.98	2.45	1.09 (3.53)
L	–0.17	–0.64	–4.33	–1.20	0.58	–0.82	–0.98	3.66	–0.49 (2.21)
M	–0.25	–1.69	6.70	–0.33	–5.43	–5.65	–2.16	0.67	–1.02 (3.89)
N	–0.26	2.68	–2.74	1.33	1.97	–0.76	2.73	2.56	0.94 (2.00)
O	–0.54	0.89	4.97	–0.33	7.63	5.82	–2.56	–0.04	1.98 (3.65)
P	–0.74	–2.97	–1.15	–1.38	–2.22	–5.04	1.24	5.99	–0.78 (3.28)
Q	–1.38	1.48	3.69	0.23	0.67	–0.42	0.54	–5.08	–0.03 (2.52)
R	–1.41	–2.23	3.49	0.50	–2.47	4.15	–9.02	–4.45	–1.43 (4.27)
S	–1.57	–5.18	–1.51	–2.20	0.26	4.79	–0.47	–1.92	–0.98 (2.82)
T	–2.11	–0.91	–1.76	–0.23	2.25	3.07	0.43	3.20	0.49 (2.12)
U	–2.84	–1.34	–0.59	–0.13	–1.34	–4.64	–0.48	0.65	–1.34 (1.69)
V	–3.08	–6.40	–0.54	–0.01	–1.09	–1.48	–0.54	1.66	–1.44 (2.41)
Mean (SD)	–0.02 (1.89)	0.00 (3.45)	–0.09 (3.38)	0.05 (1.15)	0.24 (2.65)	0.10 (3.23)	0.17 (3.35)	0.13 (3.68)	NA

Abbreviations: NA, not applicable; NICU, neonatal intensive care unit.

[a] A *z* score of zero indicates that the observed results on the measures of quality of care equal the expected results. A positive number indicates that performance is better than expected. Substantial within-NICU and between-NICU variation exists.

Table 4. Correlations Among Measures of Quality of Care, Showing That Most Are Weak[a]

No.	Measure	1	2	3	4	5	6	7	8
1	Survival	1 [Reference]							
2	Any antenatal corticosteroid use	0.42[b]	1 [Reference]						
3	No hypothermia at 1 h of life	–0.06	0.01	1 [Reference]					
4	No pneumothorax	0.38[b]	0.43[c]	0.02	1 [Reference]				
5	No health care–associated infection	0.05	0.09	0.01	–0.01	1 [Reference]			
6	High growth velocity	0.08	–0.07	0.03	0.05	0.61[c]	1 [Reference]		
7	No chronic lung disease at 36 wk gestational age	0.23	0.46[c]	–0.33	–0.04	–0.07	–0.41[b]	1 [Reference]	
8	Discharge on any human breast milk	–0.23	–0.05	–0.16	–0.18	0.11	–0.44[c]	0.35	1 [Reference]

[a] A significant positive coefficient indicates that neonatal intensive care units correspond in their performance on a pair of measures of quality of care. Only 6 of 28 pairwise correlations were statistically significant.

[b] $P < .10$.

[c] $P < .05$.

forts in health care and is based on few measures of quality and a handful of diseases (eg, diabetes mellitus, heart disease, and hypertension).

Quality improvement efforts may need to focus on multidimensional improvement and build more tightly connected systems of care so that performance can be raised on several measures of quality simultaneously. We believe that exploratory factor analysis yielded results that have a meaningful clinical interpretation and may help inform a multidimensional conceptual model for measuring and understanding NICU quality. These findings could be the focus of future improvement efforts based on underlying aspects of quality that have a causal effect on the outcomes and might need to be considered together in improving overall performance. Our group is testing whether the Baby-MONITOR, if designed according to this model, better predicts other quality-related constructs, such as safety culture. If repeated elsewhere, this could lead to a more parsimonious set of measures of quality of care to assess overall

NICU quality and offer a welcome relief to those who have to collect them. However, the limitation of such a data-driven approach may be that it could exclude the wisdom of the clinical community.

Our findings are open to different interpretations. Measures of quality may be functionally independent from each other. However, based on the clinical literature, we would a priori have expected stronger correlations. Many of the measures of quality of care (such as mortality in the NICU and antenatal corticosteroid use) have demonstrated strong causal links in randomized controlled trials.[29] We speculate that providers who excelled in one area of quality would similarly excel in others; furthermore, NICUs that reliably followed processes to avoid health care–associated infections would achieve better growth velocity and lower mortality.

We interpret our results to spotlight a low degree of systems integration within the NICU setting. Neonatal intensive care may not exist as a tightly integrated and standardized care delivery system. The NICUs seem to have the ability to excel in some areas of care but not in others.

One way to promote systems-based care may be to meaningfully measure overall quality of care by combining individual measures of quality into a composite indicator of quality.[21] One study[30] showed that, while adherence to individual process measures of surgical infection prevention did not predict postoperative infection, a composite of prevention measures did; similarly, the study found a quality signal based on the sum of ranks across NICUs, which would have been difficult to detect based on individual measures of quality. Our group is working to develop a composite indicator of NICU quality, the Baby-MONITOR, based on an explicit and rigorous framework.[14] Until such composite indicators have been developed and tested in a rigorous manner to ensure internal validity and external validity, it seems that conclusions about overall quality of care based on measurement of restricted measure sets should be viewed with skepticism.

This study must be evaluated within the context of its design. Our investigation relies on data submitted to the CPQCC by the NICUs and not by independent medical records abstractors. This may raise concern regarding the validity of the data. However, little incentive exists for NICUs to systematically submit inaccurate data because this would diminish the usefulness of

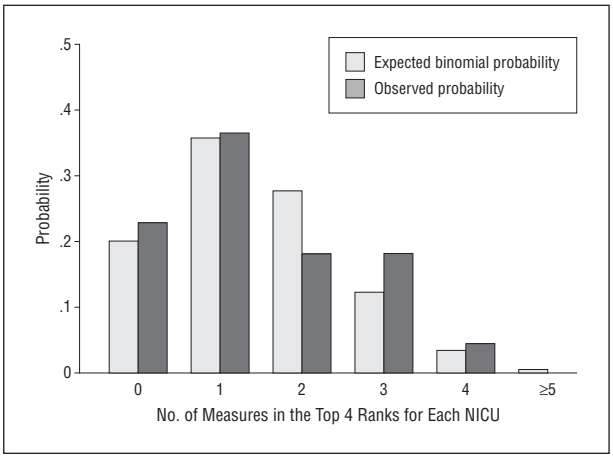

Figure 1. Observed and expected probability of a neonatal intensive care unit (NICU) being in the top 4 ranks for each measure was not different from random variation (random binomial distribution). This indicates that NICUs are not consistent with regard to performance across 8 measures of quality. If NICUs that performed in the top 4 on one measure of quality also were more likely to perform well on other measures, we would see a U-shaped distribution.

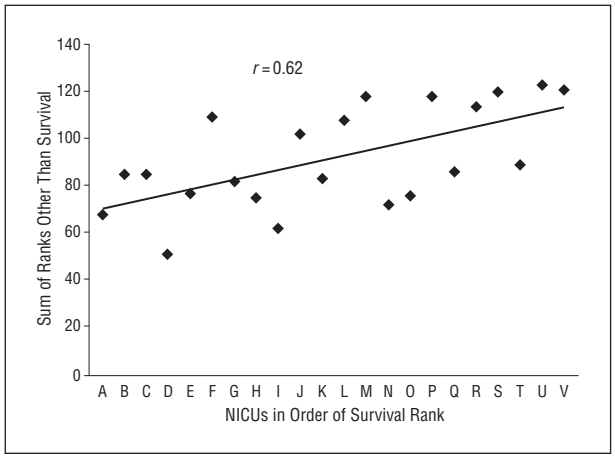

Figure 2. Correlation between survival rank and the sum of neonatal intensive care unit (NICU) ranks across 7 measures of quality of care shows a trend indicating that NICUs performing well on survival tend to perform well on the other measures. This trend was not apparent in correlations between pairs of measures of quality of care and suggests that composite measurement may provide a better global picture of quality of care delivery.

Table 5. Exploratory Factor Analysis of Performance Measures, Yielding 4 Factors

	Factor[a]			
Measure	Perinatal Care	Healthy Development	Respiratory Care	Maternal Involvement
Antenatal corticosteroid use	**0.71**	0.03	0.39	0.03
Pneumothorax	**0.67**	0.00	–0.09	–0.14
Mortality in the NICU	**0.50**	0.07	0.24	–0.24
Health care–associated infection	0.07	**0.72**	0.01	0.22
Growth velocity	–0.01	**0.94**	–0.20	–0.28
Chronic lung disease at 36 wk gestational age	0.12	–0.18	**0.96**	0.18
Discharge on any human breast milk	–0.05	–0.14	0.16	**0.97**
Hypothermia during the first hour of life	0.10	–0.07	–0.35	–0.11

Abbreviation: NICU, neonatal intensive care unit.

[a] Promax-rotated factor loadings for the 4-factor solution. Promax is a nonorthogonal rotation method used in factor analysis to enhance the interpretability of the factor-loading structure. Boldfaced terms uniquely loaded on a single factor.

data feedback from the CPQCC, a service that NICUs pay for. In addition, data validity is strengthened by the CPQCC's use of standardized data abstraction protocols and operation manuals, as well as by automated data quality management tools to identify potentially inaccurate data entries.

An alternative explanation for our findings of modest correlation of NICU performance across different measures of quality could be that quality of care among our small sample of California regional NICUs was similar. It may be that specific state-level policies foster care processes and cultures that are alike, making it harder to find diverging performance. On the other hand, investigations have found large differences in performance across other networks.[31] Nevertheless, the specific attributes of the present study may hamper generalizability to other states and types of NICUs.

We developed individual risk adjustment models to control for confounding due to clinical risk at birth. These models have not been validated in other samples; therefore, it is possible that the models introduced bias into our results, although the direction of this bias is not easily ascertained. Similarly, residual confounding introduced by unobserved variables (such as academic affiliation or staffing ratios) may have influenced our results.

In conclusion, modest correlations of NICU performance on multiple measures of quality were observed. Benchmarking of NICU quality based on isolated indicators of quality may not reflect or improve overall quality of care. Multidimensional measurement of performance via composite indicators might promote multidimensional improvement using systems-based interventions.

Accepted for Publication: May 31, 2012.
Published Online: November 12, 2012. doi:10.1001/jamapediatrics.2013.418
Author Affiliations: Department of Pediatrics (Dr Profit) and Section of Health Services Research, Department of Medicine (Drs Profit, Pietz, Hysong, and Petersen and Mr Kowalkowski), Baylor College of Medicine, Section of Neonatology, Department of Pediatrics, Texas Children's Hospital (Dr Profit), and Houston Veterans Affairs Health Services Research and Development Center of Excellence, Health Policy and Quality Program, Michael E. DeBakey Veterans Affairs Medical Center (Drs Profit, Pietz, Hysong, and Petersen and Mr Kowalkowski), Houston; Department of Neonatology, Beth Israel Deaconess Medical Center, and Division of Newborn Medicine, Harvard Medical School, Boston, Massachusetts (Dr Zupancic); and Perinatal Epidemiology and Health Outcomes Research Unit, Division of Neonatology, Stanford University School of Medicine and Lucile Packard Children's Hospital, and California Perinatal Quality Care Collaborative, Palo Alto (Dr Gould), and Department of Applied Mathematics and Statistics, Baskin School of Engineering, University of California, Santa Cruz (Dr Draper).
Correspondence: Jochen Profit, MD, MPH, Houston Veterans Affairs Health Services Research and Development Center of Excellence, Health Policy and Quality Program, Michael E. DeBakey Veterans Affairs Medical Center (152), 2002 Holcombe Blvd, Houston, TX 77030 (profit@bcm.edu).
Author Contributions: Drs Profit and Pietz had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. *Study concept and design:* Profit, Zupancic, Gould, Pietz, Draper, and Petersen. *Acquisition of data:* Profit, Zupancic, Gould, Kowalkowski, Hysong, and Petersen. *Analysis and interpretation of data:* Profit, Pietz, Kowalkowski, and Draper. *Drafting of the manuscript:* Profit, Zupancic, Gould, and Petersen. *Critical revision of the manuscript for important intellectual content:* Profit, Zupancic, Gould, Pietz, Kowalkowski, Draper, Hysong, and Petersen. *Obtained funding:* Profit, Hysong, and Petersen. *Administrative, technical, and material support:* Profit, Zupancic, Gould, and Petersen.
Conflict of Interest Disclosures: None reported.
Funding/Support: This study is supported in part by grants 1 K23 HD056298 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (Dr Profit), CD2-07-0181 from the Department of Veterans Affairs Health Services Research and Development Program (Dr Hysong), Veterans Affairs Health Services Research and Development Center of Excellence grant HFP90-20 (Drs Hysong and Petersen), and American Heart Association Established Investigator Award 0540043N (Dr Petersen).
Online-Only Material: The eAppendix is available at http://www.jamapeds.com.
Additional Contributions: We thank Aloka L. Patel, MD, and Rush University Medical Center (Chicago, Illinois) for granting Dr Profit a nonexclusive license to use Rush University Medical Center's exponential infant growth model for noncommercial research purposes.

REFERENCES

1. Institute of Medicine. *Rewarding Provider Performance: Aligning Incentives in Medicare.* Washington, DC: National Academies Press; 2006.
2. Profit J, Woodard LD. Perils and opportunities of comparative performance measurement. *Arch Pediatr Adolesc Med.* 2012;166(2):191-194.
3. Epstein AM. Paying for performance in the United States and abroad. *N Engl J Med.* 2006;355(4):406-408.
4. Rosenthal MB, Landon BE, Normand SL, Frank RG, Epstein AM. Pay for performance in commercial HMOs. *N Engl J Med.* 2006;355(18):1895-1902.
5. Petersen LA, Woodard LD, Urech T, Daw C, Sookanan S. Does pay-for-performance improve the quality of health care? *Ann Intern Med.* 2006;145(4):265-272.
6. National Quality Forum. National voluntary consensus standards for perinatal care. May 2009. http://www.qualityforum.org/Publications/2009/05/National_Voluntary_Consensus_Standards_for_Perinatal_Care_2008.aspx. Accessed September 14, 2012.
7. Shahian DM, Wolf RE, Iezzoni LI, Kirle L, Normand SL. Variability in the measurement of hospital-wide mortality rates. *N Engl J Med.* 2010;363(26):2530-2539.
8. Rosenthal GE. Weak associations between hospital mortality rates for individual diagnoses: implications for profiling hospital quality. *Am J Public Health.* 1997;87(3):429-433.
9. Wilson IB, Landon BE, Marsden PV, et al. Correlations among measures of quality in HIV care in the United States: cross sectional study. *BMJ.* 2007;335(7629):1085-1091.
10. Gandhi TK, Francis EC, Puopolo AL, Burstin HR, Haas JS, Brennan TA. Inconsistent report cards: assessing the comparability of various measures of the quality of ambulatory care. *Med Care.* 2002;40(2):155-165.
11. Palmer RH, Wright EA, Orav EJ, Hargraves JL, Louis TA. Consistency in performance among primary care practitioners. *Med Care.* 1996;34(9)(suppl):SS52-SS66.

12. Glance LGM, Osler TMM, Mukamel DBP, Dick AWP. Impact of the present-on-admission indicator on hospital quality measurement: experience with the Agency for Healthcare Research and Quality (AHRQ) Inpatient Quality Indicators. *Med Care*. 2008;46(2):112-119.
13. Ash AP. Measuring quality. *Med Care*. 2008;46(2):105-108.
14. Profit J, Typpo KV, Hysong SJ, Woodard LD, Kallen MA, Petersen LA. Improving benchmarking by using an explicit framework for the development of composite indicators: an example using pediatric quality of care. *Implement Sci*. 2010; 5:e13. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2831823/. Accessed September 14, 2012.
15. Gould JB. The role of regional collaboratives: the California Perinatal Quality Care Collaborative model. *Clin Perinatol*. 2010;37(1):71-86.
16. Horbar JD, Soll RF, Edwards WH. The Vermont Oxford Network: a community of practice. *Clin Perinatol*. 2010;37(1):29-47.
17. Profit J, Etchegaray J, Petersen LA, et al. Neonatal intensive care unit safety culture varies widely. *Arch Dis Child Fetal Neonatal Ed*. 2012;97(2):F120-F126.
18. Sexton JB, Berenholtz SM, Goeschel CA, et al. Assessing and improving safety climate in a large cohort of intensive care units. *Crit Care Med*. 2011;39(5): 934-939.
19. Pronovost PJ, Goeschel CA, Colantuoni E, et al. Sustaining reductions in catheter related bloodstream infections in Michigan intensive care units: observational study. *BMJ*. 2010;340:c309.
20. Profit J, Etchegaray J, Petersen LA, et al. The Safety Attitudes Questionnaire as a tool for benchmarking safety culture in the NICU. *Arch Dis Child Fetal Neonatal Ed*. 2012;97(2):F127-F132.
21. Profit J, Gould JB, Zupancic JA, et al. Formal selection of measures for a composite index of NICU quality of care: Baby-MONITOR. *J Perinatol*. 2011;31 (11):702-710.
22. Kowalkowski M, Gould JB, Bose C, Petersen LA, Profit J. Do practicing clinicians agree with expert ratings of neonatal intensive care unit quality measures? *J Perinatol*. 2012;32(4):247-252.
23. Stark AR; American Academy of Pediatrics Committee on Fetus and Newborn. Levels of neonatal care [published correction appears in *Pediatrics*. 2005;115(4):1118]. *Pediatrics*. 2004;114(5):1341-1347.
24. California Perinatal Quality Care Collaborative. CPQCC. http://www.cpqcc.org. Accessed February 4, 2008.
25. Patel AL, Engstrom JL, Meier PP, Kimura RE. Accuracy of methods for calculating postnatal growth velocity for extremely low birth weight infants. *Pediatrics*. 2005;116(6):1466-1473.
26. Hosmer DW, Lemeshow S. *Applied Logistic Regression*. New York, NY: John Wiley & Sons; 2001.
27. Draper D, Gittoes M. Statistical analysis of performance indicators in UK higher education. *JR Stat Soc*. 2004;167(3):449-474.
28. Aberson CL. *Applied Power Analysis for the Behavioral Sciences* . New York, NY: Taylor & Francis Group LLC; 2010.
29. NIH Consensus Development Panel on the Effect of Corticosteroids for Fetal Maturation on Perinatal Outcomes. Effect of corticosteroids for fetal maturation on perinatal outcomes. *JAMA*. 1995;273(5):413-418.
30. Stulberg JJ, Delaney CP, Neuhauser DV, Aron DC, Fu P, Koroukian SM. Adherence to surgical care improvement project measures and the association with postoperative infections. *JAMA*. 2010;303(24):2479-2485.
31. Lemons JA, Bauer CR, Oh W, et al; NICHD Neonatal Research Network. Very low birth weight outcomes of the National Institute of Child Health and Human Development Neonatal Research Network, January 1995 through December 1996. *Pediatrics*. 2001;107(1):E1.