Indirect vs Direct Hospital Quality Indicators for Very Low-Birth-Weight Infants

Jeannette A. Rogowski, PhD
Jeffrey D. Horbar, MD
Douglas O. Staiger, PhD
Michael Kenny, MS
Joseph Carpenter, MS
Jeffrey Geppert, MA

Context Evidence-based selective referral strategies are being used by an increasing number of insurers to ensure that medical care is provided by high-quality providers. In the absence of direct-quality measures based on patient outcomes, the standards currently in place for many conditions rely on indirect-quality measures such as patient volume.

Objectives To assess the potential usefulness of volume as a quality indicator for very low-birth-weight (VLBW) infants and compare volume with other potential indicators based on readily available hospital characteristics and patient outcomes.

Design, Setting, and Participants A retrospective study of 94110 VLBW infants weighing 501 to 1500 g born in 332 Vermont Oxford Network hospitals with neonatal intensive care units between January 1, 1995, and December 31, 2000.

Main Outcome Measures Mortality among VLBW infants prior to discharge home; detailed case-mix adjustment was performed by using patient characteristics available immediately after birth.

Results In hospitals with less than 50 annual admissions of VLBW infants, an additional 10 admissions were associated with an 11% reduction in mortality (95% confidence interval [CI], 5%-16%; $P<.001$). The annual volume of admissions only explained 9% of the variation across hospitals in mortality rates, and other readily available hospital characteristics explained an additional 7%. Historical volume was not significantly related to mortality rates in 1999-2000, implying that volume cannot prospectively identify high-quality providers. In contrast, hospitals in the lowest mortality quintile between 1995 and 1998 were found to have significantly lower mortality rates in 1999-2000 (odds ratio [OR], 0.64; 95% CI, 0.55-0.76; $P<.001$) and hospitals in the highest mortality quintile between 1995 and 1998 had significantly higher mortality rates in 1999-2000 (OR, 1.37; 95% CI, 1.16-1.64; $P<.001$). The percentage of hospital-level variation in mortality in 1999-2000 that was forecasted by the highest and lowest quintiles based on patient mortality was 34% compared with only 1% for the highest and lowest quintiles of volume.

Conclusions Referral of VLBW infants based on indirect-quality indicators such as patient volume may be minimally effective. Direct measures based on patient outcomes are more useful quality indicators for the purposes of selective referral, as they are better predictors of future mortality rates among providers and could save more lives.

JAMA. 2004;291:202-209 www.jama.com

THE PROCESSES AND OUTCOMES of medical care for very low-birth-weight (VLBW) infants vary markedly among different neonatal intensive care units (NICUs), even after detailed adjustment for patient risk.[1] Similar observations in a broad range of clinical specialties have led to a renewed emphasis on the accountability of health care organizations to both consumers and purchasers for the quality and safety of medical care.[2-7] For example, the Leapfrog group, a business roundtable-sponsored collaborative of large employers and insurers, has recently implemented an evidence-based referral standard for 5 surgical conditions (coronary artery bypass graft, percutaneous coronary intervention, abdominal aortic aneurysm repair, pancreatic resection, and esophagectomy) and 2 neonatal conditions (VLBW and major congenital anomalies). The Leapfrog group Web site (http://www.leapfroggroup.org) now includes hospital-specific information about compliance with these standards.

Evidence-based referral broadly means making sure that patients with high-risk conditions are treated in hospitals with the best outcomes. For some clinical conditions, such as coronary artery bypass graft surgery, in which several states publish risk-adjusted mortality rates by provider, selective referral could be based on patient-outcome indicators.[8,9] However, given the paucity of reliable publicly available risk-adjusted outcome data for most clinical procedures and conditions, the Leap-

See also p 195.

Author Affiliations: RAND, Arlington, Va (Dr Rogowski); Department of Pediatrics, University of Vermont College of Medicine and Vermont Oxford Network, Burlington (Dr Horbar); Department of Economics, Dartmouth College, Hanover, NH, and the National Bureau of Economic Research, Cambridge, Mass (Dr Staiger); Biometry Division, University of Vermont, Burlington (Mr Kenny); Vermont Oxford Network, Burlington (Mr Carpenter); and National Bureau of Economic Research, Palo Alto, Calif (Mr Geppert).

Financial Disclosures: Dr Rogowski is a member of the Vermont Oxford Network Database Advisory Committee. Dr Horbar is chief executive and scientific officer for the Vermont Oxford Network. Mr Carpenter is an employee of the Vermont Oxford Network.

Corresponding Author and Reprints: Jeannette A. Rogowski, PhD, RAND, 1200 S Hayes St, Arlington, VA 22202 (e-mail: jar@rand.org).

frog group has initially based referral decisions on hospital characteristics associated with better outcomes. For VLBW infants, the Leapfrog group's evidence-based hospital referral standard requires that infants with expected birth weight of less than 1500 g, a gestational age of less than 32 weeks, or correctable major birth defects should be delivered at a regional NICU with an average daily census of 15 or more.[10]

This article assesses the potential usefulness of indirect-quality indicators, such as volume for VLBW infants, and compares indirect-quality indicators with direct measures, such as those based on observed mortality. In particular, we address 2 related questions. First, what proportion of the variation across hospitals in mortality among VLBW infants can be explained by indirect-quality measures such as volume and NICU level? Second, can either indirect- or direct-quality indicators prospectively identify high- and low-mortality hospitals? For any referral standard to be successful, it must be able to reliably identify hospitals that will have better patient outcomes during the coming year.

METHODS

Sites and Patient Sample

The Vermont Oxford Network (VON) is a voluntary collaborative network of hospitals with NICUs located in 49 states and 22 foreign countries. Membership in the VON increased from 138 hospitals in 1995 to 353 hospitals in 2000. The VON database contains detailed uniform clinical and treatment information on all VLBW infants (<1500 g) cared for by network hospitals. By the year 2000, the VON database included approximately half of all VLBW infants born in the United States.

A total of 332 US hospitals participating in the VON provided data for this retrospective study. Of these hospitals, 108 provided data for all years and 224 provided data for 1 to 5 years and were mainly composed of those hospitals that joined the VON after 1995. The study population consisted of 94 110 infants who weighed between 501 and 1500 g born in VON hospitals between January 1, 1995, and December 31, 2000. This included infants who died in the delivery department or other locations in the hospital even if they were not admitted to the NICU. Infants born outside of the VON system were excluded from the analysis because hospital of birth has been shown to be the most important factor in infant survival.[11,12] Infants who weighed 500 g or less were excluded from the analysis for consistency with prior studies. Institutional review board approval was obtained from RAND, the University of Vermont, and the National Bureau of Economic Research.

Variables

All patient-level variables for the analyses were obtained or derived from the VON database. The key outcome measure was mortality prior to discharge home. The VON database follows infants through subsequent transfers to determine their ultimate disposition. All infant characteristics were measured at the time of birth. In addition, the median income (in $1000s) and mean education level (in years) in the ZIP code where the mother resided was derived from the 1990 census and used as an estimate of the mother's income and education. With the exception of birth weight and sex, which were available in every case, missing values for all variables were imputed. Less than 2% of infants had missing data for any given variable.

Hospital-level variables were assigned to infants based on the hospital in which the birth occurred. Volume was measured as the annual number of VLBW infants admitted to the hospital and was imputed based on the mean monthly volume for hospitals with a partial year of data. In contrast with some prior research,[11] we did not use mean daily census in the NICU as a measure of volume because the necessary data were not readily available in the VON database and because the number of VLBW infants was likely to be a more accurate indicator of a hospital's experience treating VLBW infants. The level of the NICU was derived from the VON's annual institutional survey. The VON assigns each NICU to 1 of 3 levels: level A (restriction on ventilation, minor surgery only), level B (major surgery), and level C (cardiac surgery). Levels A, B, and C correspond to high level II and level III units according to the American Academy of Pediatrics classifications of NICUs. The remaining hospital-level variables were derived from the American Hospital Association's annual survey of hospitals and the area resource file.

Statistical Models

Our focus in this study is on hospital-level determinants of mortality among VLBW infants. The analyses were based on logistic regression, with mortality prior to being discharged home as the dependent variable. The logistic regression models were estimated with only patient-level risk factors as independent variables and also estimated with both patient-level and hospital-level variables. We estimated random-effects logistic models by the method of maximum likelihood.[13] This method allows for an unobserved hospital-level component (the random effect), which captures any hospital-level factors that were omitted from the model and systematically increases or decreases mortality of all infants in that hospital. Inclusion of this random effect corrects the standard errors for the resulting within-hospital correlation (clustering) in patient outcomes and provides an estimate of the standard deviation of these unobserved differences across hospitals. We use this estimate to quantify the variance in mortality across hospitals that can be explained by volume and other hospital-level variables included in the model vs hospital-level factors that were omitted from the model. All of the substantive results reported were similar when models were estimated by using standard logit models that corrected for clustering or by using the generalized estimating equation approach.[14,15] The analyses were conducted using Stata statistical software version 7.0 (StataCorp, College Station, Tex).

To control for differences across hospitals in case mix, the logistic model included infant characteristics measured at the time of birth that are associated with mortality risk and were developed for the VON risk-adjustment model.[7] These covariates included gestational age in weeks (and its square); small for gestational age, defined by a birth weight of less than the 10th percentile for gestational age based on race and sex, derived from the 1993 US Center for Health Statistics Natality data set (NCHS, Hyattsville, Md); 1-minute Apgar score (ranging from 0-10, with higher values indicating better health); race (non-Hispanic black, non-Hispanic white, or other [including Hispanic]); sex; multiple birth; presence of a major birth defect; vaginal delivery; and whether the mother received any prenatal care.

In prior work,[7] this risk-adjustment model has been used successfully by the VON to adjust for differences in case mix for VLBW infants, with a Hosmer-Lemeshow goodness-of-fit ($P = .79$) and area under the receiver operating characteristic (AUROC) curve of 0.88. The risk-adjustment model compares well with physiologically based measures such as the Score for Neonatal Acute Physiology.[16-18] For instance, the AUROC curve for models including this score and birth weight is 0.73 for infants who weigh less than 750 g, 0.84 for infants weighing 750 to 999 g, and 0.91 for infants weighing 1000 to 1499 g. We include small for gestational age in the models to identify infants that are of low birth weight for a given gestational age and, therefore, expected to have higher mortality. Replacing small for gestational age with birth weight does not improve model fit and has no substantive impact on the results. Finally, we added to the risk adjustment model the median income (in $1000s) and mean education level (in years) from the mother's ZIP code of residence because these maternal characteristics might be confounded with the quality of the hospital in which the delivery occurred.

Prior work[11] suggested that the volume-outcome relationship was nonlinear, with lower volume only associated with higher mortality below some threshold. To investigate this relationship, we estimated a logistic model of mortality that included only patient-level covariates and used the resulting estimates to form the standardized mortality ratio for each hospital in our sample (the number of actual deaths divided by the number of expected deaths based on the logistic model). We then estimated the relationship between annual volume in each hospital and the standardized mortality ratio by using lowess regression,[19] a smoothing method that estimates the relationship between the 2 variables nonparametrically.

Based on these preliminary results, volume was allowed to have a piecewise linear effect in all of the estimated models. In particular, we estimated 3 parameters: the threshold level of volume and the effect of volume on mortality below and above the vol-

Table 1. Infant and Hospital Characteristics

	Value
Hospital-level factors (N = 332)	
Annual No. of VLBW admissions	
Mean (SD)	80.22 (51.49)
Percentile of annual VLBW admissions	
10th	25
25th	40
50th	68.5
75th	110
90th	153
NICU level, %	
A	18
B	57
C	25
Hospital in metropolitan area (>1 million residents), %	57
Hospital ownership, %	
Public	11
For-profit	8
Not-for-profit	81
Member, Council of Teaching Hospitals, %	38
Medicaid as % of admissions, mean (SD)	17 (11)
Infant-level factors (N = 94 110)	
Died before discharge home, %	14
Birth weight, mean (SD), g	1048.32 (287.30)
Gestational age, mean (SD), wk	28.46 (2.96)
1-Minute Apgar score, mean (SD)	5.39 (2.45)
Small for gestational age, %	21
Multiple birth, %	28
Congenital malformation, %	4
Mode of delivery, %	
Vaginal	38
Cesarean	62
Had prenatal care, %	96
Sex, male, %	51
Race, %	
Non-Hispanic white	56
Non-Hispanic black	28
Other*	16
Median income by ZIP code in $1000s, mean (SD)	36.19 (10.63)
Education by ZIP code, mean (SD), y	12.38 (1.17)

Abbreviations: NICU, neonatal intensive care unit; VLBW, very low birth weight.
*All other races, including Hispanic.

ume threshold. Other specifications that allowed for nonlinear effects of volume, such as including indicator variables for hospital volume deciles, yielded similar qualitative and quantitative conclusions. We report the piecewise linear specification because it captures the observed relationship in a parsimonious way and because it dominated other specifications in terms of statistical significance.

To allow for the effects of other hospital characteristics that might influence quality of care, we included the following hospital-level regressors in our models: the level of the hospital's NICU (A-C), ownership of the hospital (not-for-profit, for-profit, or public), teaching status (as indicated by membership in the Council of Teaching Hospitals), location in a major urban area, and the percentage of the hospital's patient days paid by Medicaid. Year dummies were found to be insignificant in preliminary analysis and have been excluded from the models.

Finally, to evaluate the ability of volume to prospectively identify high-mortality and low-mortality hospitals, we used data for infants born from January 1, 1995, to December 31, 1998 (265 hospitals, 52 299 infants) to rank hospitals according to their mean annual volume and used this ranking to identify hospitals in the highest- and lowest-volume quintiles. We then estimated a random effects logistic model by using the subsequent 2 years of data (1999-2000) for those hospitals remaining in the VON (252 hospitals, 36 315 infants), controlling for infant characteristics and including indicators for whether the hospital was ranked in the highest or lowest quintile but not including any other hospital-level covariates. If the estimated mortality differences in the highest and lowest quintile are large and significant, mean annual volume can prospectively identify high-mortality and low-mortality hospitals as desired. For comparison to rankings based on volume, we repeated the exercise for 2 alternative methods of prospectively ranking hospitals: one based on an index of all hospital-level characteristics, as measured by the mortality rate in 1995-1998 predicted by each hospital's volume, NICU level, and other characteristics; and the second based on each hospital's recent mortality experience, as measured by its standardized mortality ratio.

RESULTS

Volume-Outcome Relationships

The average hospital in our sample admitted approximately 80 VLBW infants annually but 25% of the hospitals admitted less than 40 VLBW infants per year, and 10% admitted less than 25 per year (TABLE 1). Hospitals in the sample generally provided a high level of care, with most having either a level B (57%) or level C (25%) NICU, and more than one third being members in the Council of Teaching Hospitals. In comparison with the universe of hospitals with NICUs in the 1997 Annual Survey of the American Hospital Association, our sample contains somewhat larger units (a mean of 27 NICU beds compared with 23 in the American Hospital Association survey) and contains disproportionately fewer of the smallest hospitals (those with <5 NICU beds). Otherwise, our sample is fairly representative of hospitals with a NICU.

Similarly, the infants in our sample are generally representative of all VLBW infants. In a comparison with VLBW in-

Table 2. Results of Random-Effect Logit Models

	Odds Ratio (95% Confidence Interval)
Hospital-level factors	
Annual No. of VLBW admissions	
Estimated threshold*	50 (41-61)
Per admission increase below threshold	0.989 (0.983-0.994)
Per admission increase above threshold	1.001 (1.000-1.002)
NICU level	
A	1.146 (0.990-1.328)
B	1.122 (1.013-1.244)
C	1.0
Large metropolitan area (>1 million residents)	1.110 (1.008-1.222)
Hospital ownership	
Public	1.015 (0.888-1.161)
For-profit	0.853 (0.711-1.023)
Not-for-profit	1.0
Member, Council of Teaching Hospitals	0.980 (0.893-1.075)
Medicaid as % of admissions	1.076 (0.754-1.536)
Infant-level factors	
Gestational age, per wk	0.054 (0.047-0.061)
Gestational age squared	1.045 (1.043-1.047)
1-Minute Apgar score per point	0.756 (0.748-0.764)
Small for gestational age	2.710 (2.497-2.942)
Multiple birth	1.278 (1.208-1.352)
Congenital malformation	19.979 (18.285-21.830)
Vaginal delivery	1.279 (1.216-1.346)
Had prenatal care	0.945 (0.849-1.051)
Sex, male	1.306 (1.245-1.370)
Race	
Non-Hispanic white	1.307 (1.226-1.395)
Non-Hispanic black	1.0
Other†	1.210 (1.116-1.312)
Median income by ZIP code per $1000s	1.000 (0.996-1.004)
Mean education by ZIP code per y	0.996 (0.963-1.029)

Abbreviations: NICU, neonatal intensive care unit; VLBW, very low birth weight.
*Estimated threshold is the annual number of VLBW admissions (50) with a 95% confidence interval of 41 to 61.
†All other races, including Hispanic.

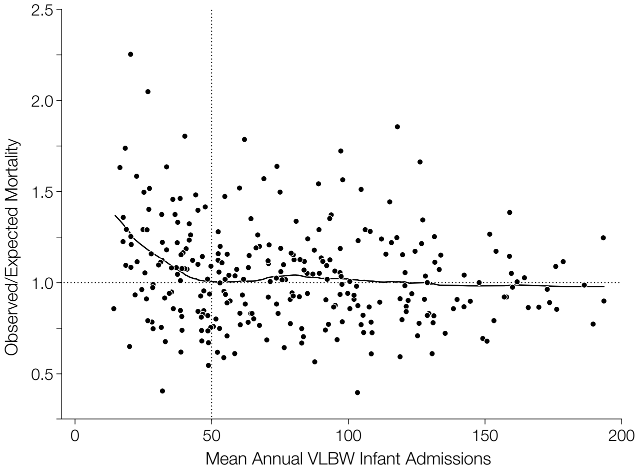
fants (500-1499 g) from the 1997 Vital Statistics data in the United States, the infants in this study have similar birth weight and gestational age distributions (a mean birth weight of 1048 g and a gestational age of 28.5 weeks in our sample vs 1040 g and 28.6 weeks in the Vital Statistics data). The race of infants in our sample was similar to the distribution among all VLBW infants from the Vital Statistics data with non-Hispanic white patients comprising 56% vs 48%, non-Hispanic black patients 28% vs 31%, and other patients (including Hispanic) 16% vs 20%, respectively. The mortality rate of infants in this study is somewhat better than among all VLBW infants (a 28-day mortality rate of 11.5% compared with 14.3%). This would be expected because the infants in our sample were all born in a hospital with a NICU, although some of this difference may be due to infants who are discharged home before 28 days and die at home, which is not recorded in the VON data.

The overall predictive power of the random-effects logit model of mortality for infants was very good, with an AUROC curve of 0.89. We found no systematic change in parameter estimates across years: the inclusion of year dummies was jointly insignificant at the 10% level and the hypothesis that all parameters were stable across the years in our sample could not be rejected at the 10% level. As a result, all 6 years of data were pooled together in the logistic analysis.

The infant-level variables have coefficients that resemble those in earlier work.[7,20] Significant effects were found for a number of infant-level variables (TABLE 2). Younger gestational age has increasingly large impacts on mortality. Taking into account the squared term, relative to a gestational age of 32 weeks, an infant born at 30 weeks is 50% more likely to die, at 28 weeks is 3 times more likely to die, and at 26 weeks is more than 9 times more likely to die. Every additional point on the Apgar score reduces mortality by approximately 25%, and being small for gestational age increases the risk of mortality by 2.7 times. Infants with major birth defects have an odds ratio (OR) of nearly 20. Controlling for all of the other risk factors, income and education of the mother's residence by ZIP code are not significantly related to mortality.

Hospital-level variables also had significant effects (Table 2). Higher volume is estimated to significantly reduce mortality (OR, 0.989; 95% confidence interval [CI], 0.983-0.994; $P<.001$) until a threshold of 50 admissions per year (95% CI, 41-61) and then slightly increase mortality (OR, 1.001; 95% CI, 1.000-1.002; $P=.045$) above this threshold. The coefficient estimates imply an 11% (95% CI, 5%-16%; $P<.001$) reduction in mortality for every additional 10 infant admissions at hospitals with less than 50 admissions per year and a 1% (95% CI, 0%-2%; $P=.045$) increase in mortality for every additional 10 infants at hospitals with more than 50 admissions per year. These estimates are quite robust across specifications and were virtually identical in specifications that did not include any other hospital-level covariates. In all specifications, we found significant volume effects below a volume threshold of around 50, and small or insignificant effects of volume above this threshold. Of the remaining hospital-level variables, being a lower-level NICU (A or B relative to C) and location in a large metropolitan area were associated with higher mortality rates of approximately 10% to 15% that were significant at or near the 5% significance level. Because differential lengths of stay may influence the in-hospital mortality measure used in this study, we tested for sensitivity to differential follow-up periods. When the analyses were repeated using 7-day and 28-day mortality, the results did not change.

Figure 1. Standardized Mortality Ratio by Mean Annual VLBW Infant Admissions to Neonatal Intensive Care Units With at Least 50 Infants Admitted Between 1995 and 2000



VLBW indicates very low-birth-weight. Solid curve is the mean standardized mortality ratio estimated with a lowess smoother, which is calculated as the ratio of observed to expected deaths in each neonatal intensive care unit between 1995 and 2000.

Although volume and other hospital characteristics are statistically significant, they explain very little of the variation in mortality across hospitals. The SD of the remaining unexplained hospital-level differences in the mortality OR is 0.34 (95% CI, 0.30-0.39; $P<.001$), implying that the typical hospital's risk-adjusted mortality rate deviates from the average by approximately plus or minus 30% to 40%. These differences are large relative to the effects of hospital-level factors that were estimated in Table 2. The proportion of the hospital-level variance in mortality that our estimates attribute to volume (9%) and other hospital-level factors (7%) is fairly small relative to the proportion attributed to the remain-

Table 3. Logit Estimates of the Effect of Alternative Hospital Rankings From 1995-1998 on Mortality in 1999-2000

	Mortality in 1999-2000, Odds Ratio (95% Confidence Interval)			
Rankings From 1995-1998 Based On	Top 20% (Lowest Expected Mortality)	Middle 60% (Referent)	Bottom 20% (Highest Expected Mortality)	Proportion of Hospital-Level Variation in Mortality in 1999-2000 Explained by Rankings
Volume alone	0.94 (0.80-1.10)	1.00	1.08 (0.86-1.36)	0.01
Volume plus other observable hospital characteristics	0.85 (0.72-1.00)	1.00	1.16 (0.93-1.44)	0.05
Standardized mortality ratio (observed/expected)	0.64 (0.55-0.76)	1.00	1.37 (1.16-1.64)	0.34

ing unexplained hospital-level effect (84%). These estimates imply that systematic differences in mortality across hospitals are large relative to the differences that can be predicted by readily observable hospital characteristics such as volume and level of the NICU.

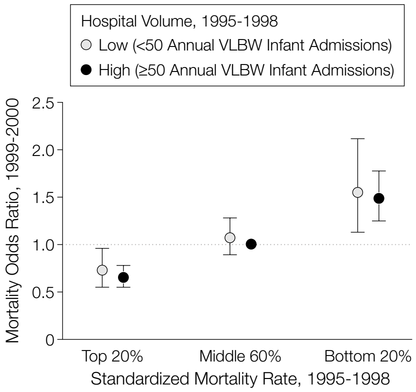
If the mean annual volume of VLBW infants is plotted against the standardized mortality ratio for every hospital with at least 50 VLBW infants (FIGURE 1), a nonlinear volume-outcome relationship similar to that estimated in Table 2 is shown, with mean mortality declining steadily at each volume level until a threshold of roughly 50 VLBW infants per year. More importantly, there are large hospital-level differences in mortality for hospitals with similar levels of volume, with mortality at many low-volume hospitals being lower than expected and mortality at many high-volume hospitals being higher than expected.

Volume as a Quality Indicator

The fact that volume and other observable hospital characteristics explain a small fraction of hospital-level variation in mortality suggests that such variables have limited use in prospectively identifying the best or worst hospitals. If hospitals are ranked based on mean annual volume in 1995-1998, hospitals in the highest-volume quintile had somewhat lower mortality in 1999-2000 (OR, 0.94; 95% CI, 0.80-1.10; *P*=.44) and hospitals in the lowest-volume quintile had somewhat higher mortality (OR, 1.08; 95% CI, 0.86-1.36; *P*=.51) compared with hospitals in the middle 3 quintiles (TABLE 3). However, these differences were not statistically significant and the volume quintiles explained only 1% of the hospital-level variation in mortality. If hospitals are ranked based on mortality rates as predicted by volume and other observable hospital characteristics from 1995-1998, the estimated mortality differences in 1999-2000 for hospitals ranked in the highest and lowest quintile were about twice as large as for rankings based on volume alone and explained 5% of the hospital-level variation in mortality. However, these differences were not statistically significant for hospitals with high-expected mortality (OR, 1.16; 95% CI, 0.93-1.44; *P*=.19) and only marginally significant for hospitals with low-expected mortality (OR, 0.85; 95% CI, 0.72-1.00; *P*=.05).

Finally, if hospitals are ranked based on their standardized mortality ratios derived from each hospital's actual infant outcomes from 1995-1998, a dramatic improvement over rankings based only on hospital characteristics such as volume and NICU level results. The rankings prospectively identified hospitals with significantly higher mortality rates (OR, 1.37; 95% CI, 1.16-1.64; *P*<.001) and hospitals with significantly lower mortality rates (OR, 0.64; 95% CI, 0.55-0.76; *P*<.001) in 1999-2000. The estimates imply a more than 2-fold difference in mortality rates between the highest-ranked and lowest-ranked hospitals, for an adjusted mortality rate of more than 19% in the worst-ranked quintile compared with less than 9% in the best-ranked quintile. The percentage of the hospital-level variation in mortality that was explained by the highest and lowest quintiles of past mortality increased to 34%.

Figure 2. Adjusted Mortality Differences (Odds Ratios) in 1999-2000 by a Hospital's Historical Volume (1995-1998) and Standardized Mortality Rate (1995-1998)



Dotted line through black circle indicates the reference group, which is hospitals averaging more than 50 very low-birth-weight (VLBW) infants per year and from the middle 3 quintiles of adjusted mortality in 1995-1998. The hospital groupings are defined as follows based on 1995-1998 data: top 20%, middle 60%, and bottom 20% are hospitals in the lowest quintile, middle 3 quintiles, and highest quintile of adjusted mortality; low-volume and high-volume hospitals are hospitals with mean annual VLBW infant admissions less than 50 and 50 or more, respectively. Error bars indicate 95% confidence intervals.

In our sample, volume appears to be a crude indicator of mortality relative to an indicator based on recent mortality rates at each hospital. This suggests that volume standards may misclassify many hospitals in a predictable manner (ie, one could prospectively identify many low-volume hospitals that were likely to have low mortality and many high-volume hospitals that were likely to have high mortality). To test this hypothesis directly, we split the sample into high-volume and low-volume hospitals (based on whether the annual number of VLBW infants exceeded 50 in 1995-1998) and identified hospitals in the highest 20%, middle 60%, or lowest 20% of each sample based on standardized mortality ratios from 1995-1998. The relative mortality rates in 1999-2000 for the high-volume and low-volume hospitals that were ranked similarly on standardized mortality ratios (eg, both top 20%) were similar in magnitude and in no case were these differences statistically significant (**FIGURE 2**). The best low-volume hospitals had significantly lower mortality than the middle group of high-volume hospitals (OR, 0.73; 95% CI, 0.55-0.96; $P = .03$) and the worst high-volume hospitals had significantly higher mortality (OR, 1.49; 95% CI, 1.25-1.76; $P < .001$).

COMMENT

Our findings that patient volume and the level of NICU care were significantly associated with mortality rates for VLBW infants are broadly consistent with the Leapfrog standard and prior studies.[11,12,23,24] Some prior studies[20,21] failed to find a volume-outcome relationship among neonates but these studies relied on small samples of infants and hospitals that may have limited the studies' power to detect an effect of volume and did not focus on identifying threshold effects.[22] Our primary finding, however, is that indirect-quality indicators such as volume and NICU level, although statistically significant, explain very little of the variation across hospitals in mortality among VLBW infants. Volume, NICU level, and other readily available hospital characteristics explained at most 16% of the variation in mortality across our sample of hospitals and forecasted at most 5% of the variation 1 to 2 years ahead. These results suggest that the current Leapfrog standard, which is based on only volume and level of care at the NICU, would tend to be an unreliable indicator of quality in our sample of hospitals (eg, many hospitals with low patient volume or low-level NICUs would have better patient outcomes and vice versa). In fact, the Leapfrog group clearly recognizes the limitations of indirect-quality measures and views the current standard as an interim solution until validated risk-adjusted patient outcome data are available.[25]

Rankings based on past mortality at each hospital outperformed rankings based on indirect-quality indicators in their ability to prospectively identify mortality differences across hospitals. In comparison with indirect-quality indicators, rankings based on past mortality forecasted far more of the hospital-level variation in mortality (34% vs 5%) and identified hospitals with larger and more statistically significant differences in mortality. In absolute terms, even past mortality cannot forecast most of the hospital-level variation in patient outcomes. This is not surprising, given prior research documenting significant variation in mortality across years, particularly for NICUs with small numbers of patients.[26] Although past mortality rates perform better than indirect-quality indicators and can on average identify quite large differences in mortality rates, they are not a particularly reliable measure of how an individual hospital will perform in the future.

Our data from the VON represent approximately 40% of the NICUs and 50% of the VLBW infants in the United States. Therefore, our results may not be representative of the general population of infants or hospitals. Although our sample is fairly representative of the general population, some important differences must be mentioned. Our analysis is limited to infants born in hospitals with a NICU and therefore does not capture any of the variation in outcomes between hospitals with and without a NICU. In addition, the sample of VON hospitals contains disproportionately fewer of the smallest hospitals in which mortality may be particularly high. Hospitals in the VON participate in ongoing quality improvement activities that may reduce the variation in patient outcomes across hospitals. The strength of indirect-quality indicators such as volume and NICU level may differ in the general population of hospitals.

Our results suggest that direct-quality indicators based on patient mortality are likely to outperform indirect-quality indicators such as patient volume and more lives could potentially be saved if patient referrals were based on the former rather than the latter. The difference in mortality between the best and worst hospitals was more than 5 times larger when ranking hospitals on past mortality rates compared with ranking hospitals on past volume. Thus, moving patients out of hospitals with high past mortality and into hospitals with lower past mortality will have a larger impact than moving patients from low-volume to high-volume hospitals. In addition, strategies that move patients from low-volume to high-volume hospitals affect relatively few patients because low-volume hospitals treat relatively few patients. Based on our estimates, a referral strategy that moved all infants out of the lowest-ranked 20% and into the middle 60% of VON hospitals in 1999-2000 would result in 11 lives saved annually based on a historical volume standard compared with 115 annually based on historical mortality experience, a 10-fold difference.

Evaluating any actual referral strategy, of course, would be far more complicated. An actual referral strategy would be limited to pregnant women known to be at risk and would be limited to geographic areas in which there was a choice of providers. One could not expect to move all infants out of the highest-mortality hospitals, leading to fewer lives saved. On the other hand, such a standard would presumably be applied to infants outside of the VON, leading to more lives saved. Moving patients among providers may generate tradeoffs in terms of family disruptions because of care received further from home and potentially increase treatment costs. Much research remains to be performed to understand how evidence-based referral strategies can be best designed to effectively manage these tradeoffs.

Author Contributions: Dr Rogowski 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: Rogowski, Horbar, Staiger, Carpenter.

Acquisition of data: Carpenter.
Analysis and interpretation of data: Rogowski, Staiger, Kenny, Carpenter, Geppert.
Drafting of the manuscript: Rogowski, Staiger.
Critical revision of the manuscript for important intellectual content: Rogowski, Horbar, Staiger, Kenny, Carpenter, Geppert.
Statistical expertise: Staiger, Kenny, Carpenter, Geppert.
Obtained funding: Rogowski, Horbar, Staiger.
Administrative, technical, or material support: Rogowski, Horbar, Staiger.
Study supervision: Rogowski.

Funding/Support: This study was supported by grants R01 HS13371 and R01 HS10328 from the Agency for Healthcare Research and Quality.

Role of the Sponsor: The Agency for Healthcare Research and Quality did not participate in the design and conduct of the study, in the collection, analysis, and interpretation of the data, or in the preparation, review, or approval of the manuscript.

Acknowledgment: We thank the 332 institutions that participated in the Vermont Oxford Network database, without whose work this research would not have been possible.

REFERENCES

1. Horbar JD, Lucey JF. Evaluation of neonatal intensive care technologies. *Future Child.* 1995;5:139-161.
2. Marshall MN, Shekelle PG, Leatherman S, Brook RH. The public release of performance data: what do we expect to gain? a review of the evidence. *JAMA.* 2000;283:1866-1874.
3. Davies HT, Marshall MN. Public discourse of performance data: does the public get what the public wants? *Lancet.* 1999;353:1639-1640.
4. Mennemeyer ST, Morrisey MA, Howard LZ. Death and reputation: how consumers acted upon HCFA mortality information. *Inquiry.* 1997;34:117-128.
5. *HCUP Quality Indicators Version 1*. Rockville, Md: Agency for Healthcare Research and Quality; 2001.
6. Harris C. England introduces star system for hospital trusts. *BMJ.* 2001;323:709.
7. Horbar JD. The Vermont Oxford Network: evidence-based quality improvement for neonatology. *Pediatrics.* 1999;103(suppl E):350-359.
8. Hannan EL, Kilburn H Jr, Racz M, Shields E, Chassin MR. Improving the outcomes of coronary artery bypass surgery in New York State. *JAMA.* 1994;271:761-766.
9. Bentley JM, Nash DB. How pennsylvania hospitals have responded to publicly released reports on coronary artery bypass graft surgery. *Jt Comm J Qual Improv.* 1998;24:40-49.
10. The Leapfrog Group. Survey Results, Explanation of HER conditions and procedures. Available at: http://www.leapfroggroup.org/consumer_HER.htm. Accessibility verified December 11, 2003.
11. 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;276:1054-1059.
12. Cifuentes J, Bronstein J, Phibbs CS, et al. Mortality in low birth weight infants according to level of neonatal care at hospital of birth. *Pediatrics.* 2002;109:745-751.
13. Conway M. A random effects model for binary data. *Biometrics.* 1990;46:317-328.
14. Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. *Biometrics.* 1988;44:1049-1060.
15. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrics.* 1986;73:13-22.
16. Richardson D, Tarnow-Mordi WO, Lee SK. Risk adjustment for quality improvement. *Pediatrics.* 1999;103(suppl E):255-265.
17. Richardson DK, Gray JE, McCormick MC, Workman K, Goldmann DA. Score for Neonatal Acute Physiology: a physiologic severity index for neonatal intensive care. *Pediatrics.* 1993;91:617-623.
18. Richardson DK, Phibbs CS, Gray JE, et al. Birth weight and illness severity: independent predictors of neonatal mortality. *Pediatrics.* 1993;91:969-975.
19. Cleveland WS. *Visualizing Data*. Summit, NJ: Hobart Press; 1993.
20. Horbar JD, Badger GJ, Lewit EM, Rogowski J, Shiono PH. Hospital and patient characteristics associated with variation in 28-day mortality rates for very low birth weight infants. *Pediatrics.* 1997;99:149-156.
21. Tucker J, UK Neonatal Staffing Study Group. Patient volume, staffing, and workload in relation to risk-adjusted outcomes in a random stratified sample of UK neonatal intensive care units: a prospective evaluation. *Lancet.* 2002;359:99-107.
22. Pollack MM, Patel KM. Need for shift in focus in research into quality of intensive care. *Lancet.* 2002;359:95-96.
23. The International Neonatal Network. The CRIB (clinical risk index for babies) score: a tool for assessing initial neonatal risk and comparing performance of neonatal intensive care units. *Lancet.* 1993;342:193-198.
24. Paneth N, Kiely JL, Wallenstein S, Susser M. The choice of place of delivery: effect of hospital level on mortality in all singleton births in New York City. *AJDC.* 1987;141:60-64.
25. The Leapfrog Group. Available at: http://www.leapfroggroup.org/consumerintro2.htm. Accessibility verified December 11, 2003.
26. Parry GJ, Gould CR, McCabe CJ, Tarnow-Mordi WO, for International Neonatal Network and the Scottish Neonatal Consultants and Nurses Collaborative Study Group. Annual league tables of mortality in neonatal intensive care units: longitudinal study. *BMJ.* 1998;316:1931-1935.

> Humanity never stands still; it advances or retreats.
> —Louis Auguste Blanqui (1805-1881)