Supplementary Online Content


eAppendix. eMethods

This supplementary material has been provided by the authors to give readers additional information about their work.
eAppendix. eMethods

Risk Adjustment

The 8 dichotomous quality measures used in the article are survival, delivery of antenatal steroids, absence of hypothermia, absence of pneumothorax, absence of nosocomial infection, high growth rate, absence of chronic lung disease, and discharge on human milk. To assess quality of NICUs, risk adjustment variables are computed to account for the fact that factors beyond the control of the NICUs may legitimately cause differences in performance that are not due to quality problems. Candidate risk-adjustment variables were administration of prenatal care, gestational age, small for gestational age, 5-minute Apgar score, multiple birth, outborn (ie, born in another facility and transferred to the study NICU), Cesarean section, and male sex. The candidate risk adjustment variables vary by outcome measure. For example, since administration of antenatal steroids occurs before birth, only prenatal care is a legitimate risk-adjustment variable for this outcome.

The procedure we used for finalizing the list of variables is described by Hosmer and Lemeshow.1 The candidate variables for each measure were evaluated individually in bivariate analysis using logistic regression. All variables with P < .25 were candidates for inclusion in the final model. The candidate variables were added to the model one by one, and a likelihood-ratio test was used to test for inclusion in the final model. A $\chi^2$ test with .05 significance level was used for this test. The candidate list and final list of risk adjustment variables are provided below.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Candidate risk adjustment variables</th>
<th>Final risk adjustment variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antenatal steroids</td>
<td>Prenatal care, gestational age, 5-minute Apgar score, multiple birth, male, outborn, Cesarean section</td>
<td>Gestational age, small for gestational age, 5-minute Apgar score, outborn</td>
</tr>
<tr>
<td>Hypothermia*</td>
<td>Gestational age, small for gestational age, 5-minute Apgar score, male, outborn, Cesarean section, prenatal care</td>
<td>Gestational age, small for gestational age, 5-minute Apgar score, male, Cesarean section</td>
</tr>
<tr>
<td>Pneumothorax*</td>
<td>Gestational age, small for gestational age, 5-minute Apgar score, multiple birth, male, Cesarean section, prenatal care</td>
<td>Gestational age, small for gestational age, 5-minute Apgar score, multiple birth, male, Cesarean section</td>
</tr>
<tr>
<td>Nosocomial infection*</td>
<td>Gestational age, small for gestational age, 5-minute Apgar score, multiple birth, male, outborn, prenatal care</td>
<td>Gestational age, small for gestational age, 5-minute Apgar score, outborn</td>
</tr>
<tr>
<td>Chronic lung disease*</td>
<td>Gestational age, small for gestational age, 5-minute Apgar score, multiple birth, male, outborn, Cesarean section, prenatal care</td>
<td>Gestational age, small for gestational age, 5-minute Apgar score, male, outborn</td>
</tr>
<tr>
<td>Discharge on human milk</td>
<td>Gestational age, small for gestational age, 5-minute Apgar score, multiple birth, male, outborn, Cesarean section, prenatal care</td>
<td>Gestational age, 5-minute Apgar score, multiple birth, outborn, prenatal care</td>
</tr>
<tr>
<td>High growth rate</td>
<td>Gestational age, small for gestational age, 5-minute Apgar score, multiple birth, male, outborn, Cesarean section, prenatal care</td>
<td>Gestational age, small for gestational age, multiple birth, male, outborn, Cesarean section, prenatal care</td>
</tr>
<tr>
<td>Survival to discharge</td>
<td>Gestational age, small for gestational age, 5-minute Apgar score, male, outborn, Cesarean section, prenatal care</td>
<td>Gestational age, small for gestational age, 5-minute Apgar score, male, outborn</td>
</tr>
</tbody>
</table>

*These variables were reverse coded to bring all outcomes into alignment for comparative performance measurement.

No collinearity or significant interactions were found in the final model. We did not consider overfitting because our goal was to see what the data would tell us about this particular group of NICUs in the period under study. No claim is made that these NICUs are a random sample of a larger population of NICUs. If, as we found, rating the NICUs on the basis of one measure did not provide a complete picture of quality, then our results would provide a counterexample to the validity of this practice.
The Draper-Gittoes Algorithm (D-G)

The D-G algorithm was developed to evaluate the performance of institutions of higher learning in the UK. It has been validated and is in current use in that arena. The algorithm requires that continuous variables be categorized. The intersection of all risk adjustment categories is computed to divide the total patient population into disjointed cells. For each measure, the observed value for an NICU is the rate of compliance for that NICU with its patients. The expected value for each risk adjustment cell is the rate of compliance in that cell in the overall database. The expected value for each NICU is a weighted average of the cell-expected values according to the number of patients the NICU has in each cell.

z Scores

The basis for evaluating NICUs is their observed minus expected rate (O-E). This number can be regarded as an estimate of an NICU’s performance if it had a very large number of patients. If the NICU had average performance, the O-E would be zero. Since only a finite number of patients are available for analysis, the estimate of O-E is not perfect, but subject to error. The D-G algorithm computes a standard error (SE) for the value of O-E for each NICU using binomial theory. The z score for each NICU is equal to (O-E)/SE. A z score is therefore a measure of an NICU’s performance that is adjusted for error. Confidence intervals are not meaningful for z scores because they already take into account the estimated error in the estimate. Since z scores are all on the same scale, a correlation matrix can be computed and factor analysis, described below, can be done.

Random-effects Models

In the D-G algorithm, the NICUs enter the analysis as fixed effects (ie, the performance of the hospital is treated as an unknown parameter to be estimated). Another type of model frequently used is a random-effects model. A random-effects model can be used to test whether a significant part of the variation between patients is due to variation between NICUs. We used random-effects models with all of the measures and the risk adjustment variables listed above as a backup analysis.

Factor Analysis

Factor analysis is based on the idea that an individual’s performance on certain measures can be explained by one or more underlying factors. It was developed as a means of explaining a student’s grades in different subjects by an underlying factor, intelligence (which cannot be measured directly). An exploratory factor analysis, used for this article, attempts to find the underlying factor structure.

Following the example of test scores and intelligence, in medical quality studies we attempt to estimate an underlying unmeasurable quantity that we think of as quality of care using scores on certain quality measures. However, the factor analysis can yield multiple factors rather than just one, as it does in this case. The factors are uncorrelated with each other.

The computation begins with the correlation matrix and produces factors that explain the correlations. The computation involves several steps in matrix algebra and produces, for each factor, numbers between −1 and 1 called factor loadings that indicate the amount of variation that is explained by that factor. Factor loadings in excess of 0.5 are considered to indicate that a significant amount of the variation in the variable is explained by the factor. The data in the study led to the 4 factors depicted in Table 5.

Reference List