Original Investigation

A Risk-Adjusted, Composite Outcomes Score and Resource Utilization Metrics for Very Low-Birth-Weight Infants

Joseph W. Kaempf, MD; John A. F. Zupancic, MD, ScD; Lian Wang, MS; Gary L. Grunkemeier, PhD

IMPORTANCE It is difficult for neonatal intensive care units (NICUs) to determine the overall efficacy of multiple continuous quality improvement (CQI) projects aimed at reducing very low-birth-weight (VLBW) infant morbidities. It is challenging to know whether a NICU is becoming more proficient, and it is not usually apparent whether concurrent resource use is changing.

OBJECTIVE To develop a risk-adjusted composite score of the major morbidities in VLBW infants and a companion metric that accounts for resource use to enhance the ability to measure overall progress in CQI and to identify proficient NICUs.

DESIGN, SETTING, AND PARTICIPANTS This retrospective investigation used individual patient-level demographic and outcomes data from 8 NICUs who were long term CQI collaborators within the Vermont Oxford Network, a large international quality improvement organization dedicated to improving the care of premature infants. Study participants were infants who weighed 401 to 1500 g born from January 1, 2000, through December 31, 2011, at each of the 8 participating NICUs.

MAIN OUTCOMES AND MEASURES Risk-adjusted, composite VLBW infant morbidity and resource utilization score.

RESULTS A total of 15,961 infants (mean [SD] gestational age, 28.2 [3.0] weeks; mean [SD] birth weight, 1020 [306] g) were analyzed. Concurrent with multiple shared CQI projects over 12 years, the group benefit metric improved 38% from 80 in 2000 to 110 in 2011 (P < .001). The entire member VON benefit metric improved 28% from 72 in 2000 to 92 in 2011 (P < .001). The group value metric improved 25% from 1.2 in 2000 to 1.5 in 2011 (P < .001). The entire member VON value metric improved 18% from 1.1 in 2000 to 1.3 in 2011 (P < .001). Significant inter-NICU variation in both composite scores was noted in the 8 member CQI group. Hospital length of stay increased in the 8 NICUs 64 to 71 days (P < .001), and a similar increase was noted in the entire member VON, 65 to 68 days (P < .001).

CONCLUSIONS AND RELEVANCE We have created the first, to our knowledge, web-based tool for NICUs to calculate their own composite morbidity and resource utilization scores that estimate NICU CQI proficiency. In our structured group CQI over 12 years, both metrics revealed significant improvement, but increases in length of stay (resource use) blunted value improvement. Why some NICUs improve their scores more successfully than others remains a crucial challenge. Future CQI efforts should explore strategies that cost-efficiently reduce intertwined VLBW infant morbidities, emphasizing whole cultures of proficient care rather than the traditional emphasis on single-morbidity reduction.

Published online March 16, 2015.

Copyright 2015 American Medical Association. All rights reserved.
T he irony of continuous quality improvement (CQI) in neonatology is that it is seldom all 3, at least by the yardstick of publication. First, it is not often truly continuous because most neonatal intensive care unit (NICU) publications that document improvement in a given morbidity do so for a circumscribed period without follow-up of sustained improvement,1-4 with notable exceptions.5 Second, the quality of specific improvement projects (assuming quality implies value—a desired health benefit achieved at a measured cost) is rarely made apparent in publications.6-7 Third, straightforward documentation of a concurrent decrease in 2 or more distinct morbidity rates or an improving composite morbidity score is remarkably rare.8

Practice changes intended to lessen one particular morbidity can be associated with an unintended increase in another morbidity,9-12 and neonologists do not often describe how cost is imbedded in CQI projects.13 How does a NICU implement and sustain evidence-based practices that improve the overall health of premature infants while also using resources more efficiently?

We propose that a reasonable representation of the proficiency of CQI projects in the NICU for very low-birth-weight (VLBW) infants might be the sum of the 8 major morbidities divided by the total hospital length of stay (and appropriately risk adjusted). A straightforward, ready-to-calculate benefit and value metric could improve CQI efforts by providing a straightforward NICU production measure that reflects overall morbidity reduction efforts. We describe the development of our composite outcomes scores and provide an example spanning 12 years as to what these measurements look like in 8 NICUs that are long-term participants in collaborative CQI.14

Methods

Our investigation used individual patient-level data from 8 NICUs within the Vermont Oxford Network (VON), a large, international quality improvement organization dedicated to improving the care of premature infants.14 Member NICUs in the VON submit standardized deidentified data for infants weighing 401 to 1500 g at birth and receive confidential reports comparing their own practices and outcomes with the network as a whole for use in quality improvement. Authorized users at member hospitals have access to their own NICU patient-level data and to aggregate data for the entire network through the VON Nightingale Internet Reporting System. An authorized individual at each participating hospital downloaded their deidentified patient-level data from the VON Nightingale Internet Reporting System and provided it to one of the authors (J.W.K.) and the statisticians for use in this study. The aggregate data in this investigation. The VON did not provide any individual hospital- or patient-level data for this study and played no role in the design, conduct, analysis, interpretation, or reporting of this research. Specific written permission for the use of the individual deidentified patient-level data was obtained from the institutional review boards of the 8 participating hospitals.

Eligible infants included all live births and admissions weighing 401 to 1500 g who were entered into the VON database from January 1, 2000, through December 31, 2011, at each of the 8 participating NICUs (listed in the Additional Contributions section). This 12-year period was selected as inclusive of our various group quality improvement projects, which began as part of the VON CQI collaborative activities.14 Our 8 NICUs (informally called the POD) formed a subgroup of the VON CQI efforts and convened regularly at biannual meetings supplemented with monthly telephone conferences. All 8 POD NICUs used various evidence-based improvement projects during the 12-year study period; although there was significant overlap, not all projects were identical at each NICU. These projects included potentially better practice (PBP) applications related to resuscitation and delivery room management, chronic lung disease, any late infection, necrotizing enterocolitis, and retinopathy of prematurity, as well as protocols to enhance breast milk use, optimal growth and nutrition, brain development, and neuroprotection. The relevant descriptive data of our POD VLBW infants are listed in the Table along with the entire concurrent VON VLBW population. The concept and development of this composite morbidity and value score were the idea of the 4 coauthors. The study purpose, methods, results, and discussion were reviewed by the coauthors along with the entire POD group for thorough feedback before submission of the manuscript.

Figure 1 shows the POD formulation of the value metric as a straightforward ratio of benefit to cost multiplied by a risk adjuster derived from the VON demographic factors listed in the Table and a mortality deduction factor. To avoid bias associated with focusing on limited or select outcomes, we defined benefit as the complement of the sum of all 8 major morbidity rates as defined by the VON manual of operations (chronic lung disease, grade 3-4 intraventricular hemorrhage, periventricular leukomalacia, stage 3-4 retinopathy of prematurity, any late infection, necrotizing enterocolitis, focal intestinal perforation, and discharge weight <10th percentile). The VON NICUs follow standardized definitions for grade 3 to 4 intraventricular hemorrhage, periventricular leukomalacia, stage 3 to 4 retinopathy of prematurity, necrotizing enterocolitis, and spontaneous intestinal perforation.15 Chronic lung disease is defined in those infants who receive supplemental oxygen at a postmenstrual age of 36 weeks or who were discharged with supplemental oxygen at gestational ages between 34 and 36 weeks. Any late infection is defined in those infants with any bacterial, coagulase-negative staphylococcal, or fungal infection after day 3 of life. Discharge weight less than the 10th percentile is defined as those infants with no major birth defect who were discharged home at gestational ages between 15 and 367 days at postmenstrual ages of 22 to 50 weeks whose weight at discharge was below Fenton’s 10th percentile for gestational age.16 The study infants may have been diagnosed as having 0, 1, or 2 or more morbidities.
We weighed all 8 major morbidities equally because any 1 morbidity or multiple permutations of the 8 have not been reported to correlate best with long-term health and/or resource expenditure.\textsuperscript{9,17} This investigation was not designed to study individual patterns of single-morbidity fluctuation; our intent was to focus on the yearly composite morbidity scores at each NICU. Total hospital length of stay is a practical measure of resource use that correlates well with overall NICU expenditures, is readily available to each NICU from their VON annual report, and does not require time-consuming tabulation of the many sources of cost and charges and thus was selected as the cost proxy.\textsuperscript{18,19}

The risk adjuster was derived in 2 steps. First, we used individual patient-level risk modeling from the study infants in the 8 POD NICUs, using the same demographic variables used by the VON in its annual report of standardized morbidity and mortality ratios (ie, gestational age, birth weight [small for gestational age], inborn, major birth defect, mode of delivery, 1-minute Apgar score, multiple birth, sex, and maternal race) (Table).\textsuperscript{20} Altitude is used by the VON as an additional risk adjuster for chronic lung disease only; we did not include this variable because all 8 POD NICUs are located below 1200-m elevation. Second, we used NICU facility-level risk modeling with simulation techniques based on predictive outcomes from the first step.

With the use of the individual patient-level data from the POD NICUs, 8 multivariable logistic regression models (1 for each of the 8 major morbidities) were produced. The corresponding predicted risks for each patient were calculated and used in the subsequent bootstrap-based simulation.\textsuperscript{21} For one simulation run, “N” separate virtual NICUs with “n” infants per unit were randomly sampled from the entire POD 12-year population. For each virtual NICU, the 8 predicted morbidity rates were computed and summed to produce the predicted composite morbidity score, and the aggregated facility-level risk factors (gestational age (<27 weeks, small for gestational age, outborn, major birth defect, cesarean section delivery, Apgar score at 1 minute <4, multiple birth, and male sex) were calculated. The relationship of the facility-level risk factors to the predicted composite morbidity score was evaluated using multivariable linear regression and was quantified through the estimated coefficient for each risk factor. A total of 6 simulation settings were used (N = 500 or 1000 and n = 100, 150, or 200) with 100 simulation runs for each setting, resulting in 600 sets of estimated coefficients from 600 linear regressions. For each risk factor, the averaged coefficient for the 600 regressions was considered to be its mean effect on the predicted composite morbidity score. For a real-world NICU, the predicted composite morbidity score was then estimated based on the facility-level risk factors using these averaged coefficients and served as the risk adjuster for the benefit and value metric.

The mortality deduction factor is calculated as 100% minus the mortality excluding early deaths (MEED) (Figure 1), which is defined as all NICU deaths excluding those that occurred in the delivery room or in the first 12 hours of life.\textsuperscript{13} This adjustment accounts for morbidities that may not be diagnosed if the infant dies (eg, chronic lung disease and retinopathy of prematurity). Mortality excluding early deaths is pref-

### Table. Demographic Profile of the POD Study Population and the Entire Concurrent VON, 2000-2011

<table>
<thead>
<tr>
<th>Group</th>
<th>No. of Very Low-Birth-Weight Infants</th>
<th>Gestational Age, Mean (SD), wk</th>
<th>Birth Weight, Mean (SD), g</th>
<th>Patients, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>NICU 1</td>
<td>1680</td>
<td>28.2 (3.0)</td>
<td>1025 (295)</td>
<td>73.5</td>
</tr>
<tr>
<td>NICU 2</td>
<td>3027</td>
<td>27.9 (3.0)</td>
<td>1002 (317)</td>
<td>84.3</td>
</tr>
<tr>
<td>NICU 3</td>
<td>1124</td>
<td>28.4 (3.0)</td>
<td>1055 (292)</td>
<td>86.6</td>
</tr>
<tr>
<td>NICU 4</td>
<td>2780</td>
<td>28.0 (3.1)</td>
<td>1009 (303)</td>
<td>85.4</td>
</tr>
<tr>
<td>NICU 5</td>
<td>1593</td>
<td>27.9 (3.2)</td>
<td>1007 (317)</td>
<td>87.6</td>
</tr>
<tr>
<td>NICU 6</td>
<td>2224</td>
<td>28.3 (2.9)</td>
<td>1029 (299)</td>
<td>94.6</td>
</tr>
<tr>
<td>NICU 7</td>
<td>1753</td>
<td>28.6 (3.0)</td>
<td>1048 (297)</td>
<td>94.0</td>
</tr>
<tr>
<td>NICU 8</td>
<td>1780</td>
<td>28.1 (3.1)</td>
<td>1012 (315)</td>
<td>98.5</td>
</tr>
<tr>
<td>POD group</td>
<td>15 961</td>
<td>28.2 (3.0)</td>
<td>1020 (306)</td>
<td>87.9</td>
</tr>
<tr>
<td>Entire VON</td>
<td>557 616</td>
<td>28</td>
<td>1042</td>
<td>84.9</td>
</tr>
</tbody>
</table>

Abbreviations: NICU, neonatal intensive care unit; POD, informal name of the participating 8 NICUs; VON, Vermont Oxford Network.

The 8 major morbidities are as follows: chronic lung disease, grade 3 to 4 intraventricular hemorrhage, periventricular leukomalacia, stage 3 to 4 retinopathy of prematurity, any late infection, necrotizing enterocolitis, focal intestinal perforation, and discharge weight less than the 10th percentile. VON indicates Vermont Oxford Network.
erable to total mortality because a significant percentage of early deaths are related to palliative care choices or multiple anomalies—events that vary among NICUs and do not necessarily reflect the quality of NICU care.\textsuperscript{8,22} From this value metric, the benefit metric is derived by omitting the denominator (mean total hospital length of stay in survivors).

Linear regression was used to examine benefit and value metrics fluctuation over time, and the $t$ test was used for comparing the mean values between the POD group and the entire VON. All data analyses and simulations were performed using R statistical software, version 2.14.0 (R Foundation for Statistical Computing).

Results

A total of 15,961 VLBW infants from the 8 POD NICUs (2000-2011) were included (mean [SD] gestational age, 28.2 [3.0] weeks; mean [SD] birth weight, 1020 [306] g). The overall demographic profile from our POD VLBW population was clinically similar to the entire VON (Table). Logistic regression analysis indicated that maternal race was not a significant risk factor across the 8 morbidities and so was not included in the facility-level risk modeling.

The formulation of the benefit metric (Figure 1) is an expected to observed ratio that is risk adjusted and multiplied by 100 (Figure 1). After deduction of the typical POD MEED rate, a mean benefit metric score would be approximately 100 − 9 = 91. Similarly, a mean value metric score would be the benefit metric divided by the typical POD length of stay in survivors (91/67 = 1.36). Our study intention was to examine the 2 metrics over time within and among NICUs and not to focus on absolute or isolated measurements. Over 12 years, the POD benefit metric composite score increased 38% from 80 in 2000 to 110 in 2011 ($P < .001$) (Figure 2). Seven of the participant NICUs had significant composite morbidity improvement, and one NICU had no change (Figure 2). The POD benefit metric was significantly greater than the VON measurement ($P = .005$). The VON benefit metric also increased significantly over 12 years, 28% from 72 to 92 ($P < .001$), but at a slower rate than that of the POD ($P < .001$) (Figure 2).

The POD value metric composite score increased 25% from 1.2 in 2000 to 1.5 in 2011 ($P < .001$) (Figure 3). Individually, 5 POD NICUs had significant improvement, 2 had no change, and 1 declined (Figure 3). The POD value metric overall was significantly greater than the VON measurement ($P = .005$). The VON value metric also increased significantly over 12 years, 18% from 1.1 to 1.3 ($P < .001$), but at a slower rate than that of the

![Figure 2. Yearly Benefit Metric Calculation in the 8 Neonatal Intensive Care Units (NICUs)](https://example.com/fig2)

![Figure 3. Yearly Value Metric Calculation in the 8 Neonatal Intensive Care Units (NICUs)](https://example.com/fig3)
POD (P = .03) (Figure 3). Total hospital length of stay in survivors significantly increased in the POD from 64 to 71 days (P < .001). This trend was also noted in the VON (65 to 68 days; P < .001) (Figure 4). The 8 individual morbidity and MEED rates for each POD NICU and the entire VON during the 12-year study period are given in eFigures 1 through 9 in the Supplement.

Discussion

From our large diverse population of VLBW infants, we have developed a risk-adjusted composite morbidity score that accounts for all 8 major NICU morbidities in VLBW infants (benefit metric) and a companion measurement that controls for the principle indicator of NICU resource use (value metric) (Figure 1). Our user-friendly CQI tool can help neonatologists formulate a big-picture evaluation of whether their NICU’s various intertwined PBPs are improving overall outcomes from year to year, a task consistent with directives from the Institute of Medicine.23

Neonatologists have adopted CQI methods, such as plan-do-study-act cycles, to facilitate improved outcomes.24 Substantial gains related to CQI projects are encouraging, but most publications4-25-27 report improvement in just one targeted morbidity and generally from an unfavorable position to an average or above average percentile. The CQI projects have several limitations: (1) single-center analyses of short duration, (2) nonrandomized design, (2) lack of clarity regarding coexisting changes in clinical care, (4) no composite score of major morbidities to assess overall health, and (5) no account of resource use. Ensuring that a NICU endorses therapies that reliably improve a broad array of outcomes and demonstrating that such PBPs can be sustained in a cost-effective fashion have thus far remained elusive, an observation noted in recent reviews of CQI.26-32 Our study is the first investigation, to our knowledge, to find significant individual and group improvement in multiple NICU morbidities that are sustained for years and also includes a measure of resource use.

Progress has been noted in a recent analysis of the outcomes of VLBW infants born in 2000 to 2009, which found improvement in a select group of morbidities and a reduction in overall mortality.33 Our composite morbidity score (benefit metric) from our 8-member CQI collaborative (and the entire VON) reveals similar improvement (Figure 2). Our composite score includes 3 additional important morbidities (retinopathy of prematurity, suboptimal NICU growth, and spontaneous intestinal perforation), and our web-based tool allows any NICU to readily estimate its own composite morbidity and resource utilization metrics.

Which PBPs, cultural factors, leadership styles, staffing structure, or as yet unmeasured variables enabled most of our POD NICUs to significantly improve their individual benefit and value metrics? Current CQI methods related to intertwined PBPs and morbidities frequently lack the requisite precision and accuracy to tease out cause and effect. Randomized clinical trials provide important insight into a given specific therapy but generally for a single morbidity and in a necessarily generalized population, leaving physicians to decide the utility of disparate PBPs.34 Profit at al8 developed a composite indicator that appears promising as a comprehensive quality measurement. Our benefit and value metric composite scores differ by including a wider variety of VLBW infant morbidities, using a distinct risk adjustment model in a larger pool of patients, and being readily accessible on our website. A comparison of these 2 composite scoring models within the same group of NICUs would be instructive.

Reducing morbidities in our POD group was associated with an unintended consequence: increased hospital length of stay (Figure 4). The length of stay in the NICU correlates strongly with resource use18,19 and the rate and severity of major VLBW morbidities.7,35,36 NICUs might find it helpful to focus on overall morbidity reduction even if such gains are associated with an increased hospital length of stay. This is an arguably acceptable trade-off financially and biologically because increased length of stay may be a reflection of increasing severity of 1 or more of the 8 morbidities and/or inefficient or wasteful care processes.

Our benefit and value metrics include all 8 major VLBW infant morbidities, providing a reasonably complete assessment of a premature infant’s health. The outcomes are weighted equally because there is no compelling evidence that any particular diagnosis, or multiple permutations of the 8, is
more (or less) harmful or expensive than the others.9,12 Inclusion of all 8 morbidities helps avoid the pitfall of selecting PBPVs that might improve one particular outcome but worsen another.9,12 We chose total hospital length of stay in survivors as a practical and meaningful proxy for overall cost because NICU bed use closely correlates with resource use.18,19 This cost estimation obviates any requirement for time-consuming tabulations of laboratory and pharmacy charges, equipment use, and professional fees. Our risk adjustment was derived from individual patient-level data from 15,961 VLBW infants in 8 diverse NICUs collected over a 12-year time span. Our VLBW demographic profile is similar to the entire VON (Table), and we used the same variables used by the VON each year to risk adjust their standardized morbidity and mortality ratios (Figure 1).

Limitations of our investigation are several. No single measurement can describe a NICU’s total proficiency or capture the elusive human elements of group intelligence and enthusiasm. It is possible that our large POD VLBW infant population from which we formulated the risk adjustment does not reflect the patient profile of a particular NICU using our web-based calculator. The 8 major morbidities included in our composite score do not account for every meaningful diagnosis in VLBW infants but were selected as the broad representation of important outcomes that the VON tracks. Our investigation was not designed to report specific patterns of individual morbidity fluctuation, but this can readily be done at the individual NICU level to enhance local CQI efforts. Our proxy for cost—total hospital length of stay in survivors—does not account for resource use related to NICU deaths and does not itemize individual expenditures. We recognize that increased length of stay may be viewed as beneficial or valuable to families in certain clinical situations. We have not proved that a higher benefit and/or value metric in the NICU necessarily improves long-term health or reduces total health care expenditures. To calculate the benefit and value metric composite scores for your NICU, we encourage readers to refer to our website featuring our CQI tool (www.providence.org/nicuvaluemetric).

Conclusions

We have created the first web-based tool for NICUs to calculate their own composite morbidity and resource utilization scores that estimate NICU CQI proficiency. In our structured group CQI work over 12 years, both metrics revealed significant improvement, but increases in length of stay (resource use) blunted value improvement. Why some NICUs improve their scores more successfully than others remains a crucial challenge. Future CQI efforts should explore strategies that cost-efficiently reduce intertwined VLBW infant morbidities, emphasizing whole cultures of proficient care rather than the traditional emphasis on single-morbidity reduction.

ARTICLE INFORMATION

Accepted for Publication: December 2, 2014.

Author Contributions: Dr Kaempf and Ms Wang 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: Kaempf, Zupancic.
Acquisition, analysis, or interpretation of data: Kaempf, Wang, Grunkemeier.
Drafting of the manuscript: Kaempf, Zupancic.
Critical revision of the manuscript for important intellectual content: Kaempf, Zupancic.
Statistical analysis: Kaempf, Wang, Grunkemeier.
Obtained funding: Kaempf.
Administrative, technical, or material support: Kaempf.
Study supervision: Kaempf.
Conflict of Interest Disclosures: None reported.
Funding/Sponsor: This investigation was supported by the Providence Health and Services Foundation and Northwest Newborn Specialists PC.
Role of the Funder/Sponsor: The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and the decision to submit the manuscript for publication.
Disclaimer: The VON played no role in the design, analysis, interpretation, or reporting of this research. The views, conclusions, and opinions expressed are solely those of the authors and do not represent those of the VON.

Additional Contributions: This research was conducted in collaboration with multidisciplinary teams from the following institutions in the United States that participated in the POD group of the VON NICQ Quality Improvement Collaborative: Children’s Hospital and Clinics of Minnesota, Minneapolis; Children’s Hospital and Clinics of Minnesota, St Paul; Helen DeVos Children's Hospital, Grand Rapids, Michigan; Lucile Packard Children's Hospital at Stanford, Palo Alto, California; Miami Valley Hospital, Dayton, Ohio; Children's Hospital at Providence Alaska, Anchorage,; Providence St. Vincent Medical Center, Portland, Oregon; and Saint Barnabas Medical Center, Livingston, New Jersey. We thank all our colleagues in the entire POD group.

REFERENCES