ARTICLE

A Cluster-Randomized Trial of Benchmarking and Multimodal Quality Improvement to Improve Rates of Survival Free of Bronchopulmonary Dysplasia for Infants With Birth Weights of Less Than 1250 Grams

Michele Walsh, MD, MS[a], Abbott Laptook, MD[b], S. Nadya Kazzi, MD, MPH[c], William A. Engle, MD[d], Qing Yao, PhD[e], Maynard Rasmussen, MD[f], Susie Buchter, MD[g], Gregory Heldt, MD[h], William Rhine, MD[i], Rose Higgins, MD[j], Kenneth Poole, PhD[e], for the National Institute of Child Health and Human Development Neonatal Research Network

[a]Department of Pediatrics, Rainbow Babies & Children's Hospital, Case Western Reserve University, Cleveland, Ohio; [b]Department of Pediatrics, Women & Infants Hospital, Brown University, Providence, Rhode Island; [c]Department of Pediatrics, Wayne State University, Detroit, Michigan; [d]Department of Pediatrics, Riley Hospital for Children, Indiana University, Indianapolis, Indiana; [e]Research Triangle Institute, Research Triangle Park, North Carolina; [f]Department of Pediatrics, Sharp Mary Birch Hospital, University of California, San Diego, San Diego, California; [g]Department of Pediatrics, Emory University, Atlanta, Georgia; [h]Department of Pediatrics, University of California, San Diego, San Diego, California; [i]Department of Pediatrics, Lucille Packard Children's Hospital, Stanford University, Palo Alto, California; [j]National Institute of Child Health and Human Development, Rockville, Maryland

The authors have indicated they have no financial relationships relevant to this article to disclose.

ABSTRACT

OBJECTIVE. We tested whether NICU teams trained in benchmarking and quality improvement would change practices and improve rates of survival without bronchopulmonary dysplasia in inborn neonates with birth weights of <1250 g.

METHODS. A cluster-randomized trial enrolled 4093 inborn neonates with birth weights of <1250 g at 17 centers of the National Institute of Child Health and Human Development Neonatal Research Network. Three centers were selected as best performers, and the remaining 14 centers were randomized to intervention or control. Changes in rates of survival free of bronchopulmonary dysplasia were compared between study year 1 and year 3.

RESULTS. Intervention centers implemented potentially better practices successfully; changes included reduced oxygen saturation targets and reduced exposure to mechanical ventilation. Five of 7 intervention centers and 2 of 7 control centers implemented use of high-saturation alarms to reduce oxygen exposure. Lower oxygen saturation targets reduced oxygen levels in the first week of life. Despite these changes, rates of survival free of bronchopulmonary dysplasia were all similar between intervention and control groups and remained significantly less than the rate achieved in the best-performing centers (73.3%).

CONCLUSIONS. In this cluster-randomized trial, benchmarking and multimodal quality improvement changed practices but did not reduce bronchopulmonary dysplasia rates.

IN 2000, PUBLICATION of the seminal report *To Err Is Human* by the Institute of Medicine rocked the medical world and shook the confidence of patients in the health care system.[1] The Institute of Medicine then examined deficiencies in common health care practices in the report *Crossing the Quality Chasm* and found that health care processes were subject to significant deviations from expected practice that led to suboptimal care and poor patient outcomes.[2] Quality improvement (QI) techniques, adapted from industry, have been used to improve care and patient outcomes. However, the National Academies of Sciences found that research on QI lagged behind real-world applications.[3] The Institute of Medicine concluded that most health care professionals were trained inadequately in QI.[3] The optimal techniques for training teams and changing practice are not known.

One technique in QI is benchmarking, in which best-performing centers are identified and practices are examined and emulated at other centers to improve outcomes. The rationale is that institutions with excellent performance for a given outcome apply specific clinical practices that are most effective. They may also display structural or cultural organizational features that contribute to excellent outcomes. By visiting these centers and reviewing the evidence in the literature, teams from other institutions can identify these practices and organizational features. Then, by applying methods learned in QI training, the teams should be able to implement the identified practices and to modify their organizations in ways that lead to better outcomes. Although such QI teams are used increasingly in health care organizations, their efficacy has not been evaluated rigorously. Evaluations of the impact of QI teams have revealed mixed results. Shojania et al[4] reviewed systematically the results of QI studies and found evidence for only modest effects, together with evidence of publication bias, with small trials being more likely to demonstrate positive results and larger trials more often yielding negative results. Also clouding the evaluation of QI is the fact that teams are often self-selected and highly motivated. It is not known whether the successful improvements in outcomes reported by such teams can be generalized when applied on a wider scale. In addition, it is unclear whether teams can select from the potentially thousands of clinical practices used to identify a subset of practices that can be applied to produce improved outcomes.

Survival rates for very low birth weight neonates (<1500 g) have improved steadily, with 83% of all such neonates surviving.[5] Although most of the survivors are healthy, many develop a chronic lung injury, bronchopulmonary dysplasia (BPD), which is a significant health burden.[6–8] In 1998, 55% of neonates with birth weights of <1250 g who were born at centers in the National Institute of Child Health and Human Development Neonatal Research Network either died or developed BPD.

The incidence rates of BPD vary by center and are not explained by differences in birth weight, gestational age, race, frequency of prenatal steroid use, or incidence of respiratory distress syndrome.[9] Therefore, differences in treatment practices may contribute to the development of BPD.[10,11] We conducted a cluster-randomized, controlled trial to test whether NICUs trained in benchmarking and multimodal QI techniques could improve rates of survival without BPD for neonates with birth weights of <1250 g, compared with centers with usual practice.

METHODS

Participants

Seventeen centers of the National Institute of Child Health and Human Development Neonatal Research Network participated in the trial, with practices analyzed for inborn neonates with birth weights of <1250 g. In January 2001, the 3 centers with the highest rates of survival free of BPD (top 3 performers in 1998–2000; rate of survival free of BPD: 62.5%) were identified as the benchmark centers (see "Acknowledgments").

Randomization

Our intention was to improve the use of potentially better practices by the entire neonatal care team. The NICU, rather than the patient, was the unit of randomization, because the intervention was applied to a team representing the NICU. In June 2001, the 14 remaining centers were assigned randomly, with computer-generated codes prepared in sealed opaque envelopes by the data center, to the intervention group ($N = 7$) or the control group ($N = 7$). Envelopes were distributed in person, and all were opened simultaneously. A flow diagram of study participants and units is shown in Fig 1.

Interventions

Data Collection

Before randomization, all 14 eligible sites selected multidisciplinary teams (neonatologist, neonatal nurse, and respiratory therapist); members were respected clinical experts at their sites. From January to June 2001, data on preintervention practices were collected at the benchmark centers and at each intervention center and were analyzed by each intervention team to identify care differences. From June to November 2001, teams conducted self-study and literature review by using the preintervention data.

Training in QI Practices

All team members attended an 8-hour training session on QI led by a team of experts (see "Acknowledgments"). Sessions introduced systems thinking, cycles of rapid change, measurement tools, and the concept of potentially better practices.[12] Teams were provided with literature reviews of care practices, including published
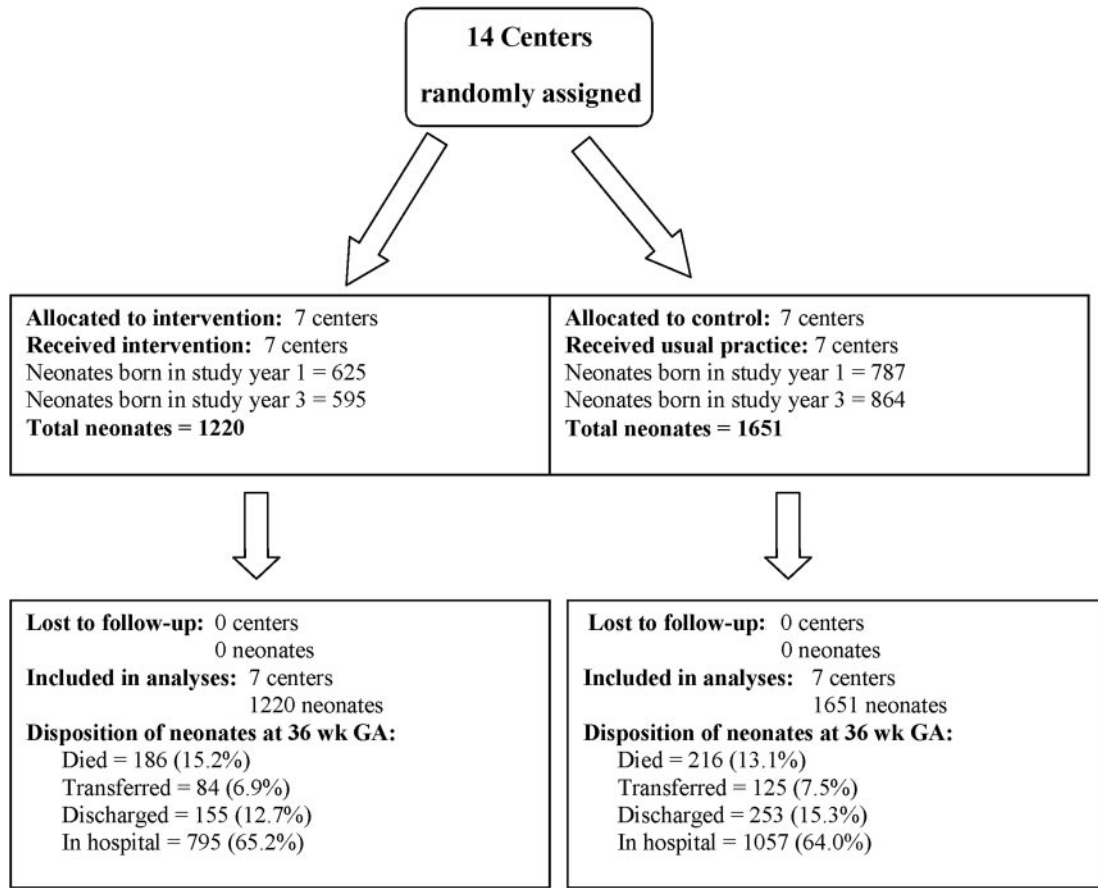
FIGURE 1
Flow diagram of participating centers and neonates in the trial. GA indicates gestational age.

meta-analyses and reviews from the Cochrane Collaboration. Core teams met face to face on 2 occasions and then in teleconferences. Teleconferences at 4- to 8-week intervals throughout the 2-year intervention period supported initial training. One control site and 2 intervention sites had participated in previous rapid-cycle QI processes.

Site Visits and Selection of Interventions
The teams from the intervention centers visited each benchmark center in November 2001. Benchmark centers delivered a presentation on their self-assessments of practices responsible for their high rates of survival free of BPD to the intervention teams. Teams also observed care directly at each benchmark site, collaborated in document care practices, and compared these with the benchmark self-assessments. In addition, intervention teams scrutinized extensive data collected by research nurses during the preintervention period at the benchmark centers and at their own centers.

From these data-driven assessments, teams identified 27 potentially better practices at the benchmark centers, in 3 domains, namely, delivery room care, ventilation practices, and nutrition and fluid practices (Table 1).

Two other domains evaluated originally, that is, infection rates and infection control practices and organizational structure, were not different at the better-performing centers and were not selected for implementation. Overall care in the benchmark centers was characterized by ventilation with lower tidal volumes (2 centers with emphasis on use of nasal continuous positive airway pressure [CPAP] and 1 center with continued mechanical ventilation with low tidal volumes) and lower oxygen saturation targets.

After the site visits, the teams reviewed published evidence, focusing on systematic reviews, evaluated the quality of the evidence by using the criteria of the Oxford Centre for Evidence-Based Medicine, and collaborated with colleagues at their centers to identify practices at the benchmark centers that were different from those in their own centers. The core team members led conferences to develop consensus with their colleagues and together selected potentially better practices for implementation. Because preexisting practices differed at each center (by design), each unit developed a unique set of interventions based on their local practice patterns; however, many centers chose similar interventions. Intervention centers chose between 5 and 13 potentially better practices per center (median: 7 practices) for im-

TABLE 1 Summary of Potentially Better Practices and Selections According to Center

Intervention	Benchmark A	Benchmark B	Benchmark C	Center A	Center B	Center C	Center D	Center E	Center F	Center G
Delivery room										
1. Fellow or attending physician at delivery	P	P	P	P		X		P	P	P
2. Respiratory therapist at delivery	P	P	P	P	P	P	P	P	P	
3. Consistent equipment in all delivery rooms	P	P	P	P	P		P	P	P	P
4. Limited pressures in manual ventilation	P	P	P	X	X	P	P		X	X
5. Prophylactic use of surfactant	P	P		P				X	X	
6. Use of device to provide positive end-expiratory pressure				P		P		P	P	
Respiratory care										
7. Selective intubation with liberal use of CPAP			P	P					X	X
8. Early use of surfactant if intubated	P	P	P	P	P	P	P		X	P
9. Assessment of volume/pressure and targeting of lowest levels to achieve modest chest rise if intubated	P	P	P	X	X		P		X	P
10. Aggressive weaning and early extubation if intubated	P	P	P	P		X	X	X	X	X
11. Higher PaCO$_2$ targets for all patients	P	P	P	P	X	P	P	P	X	X
12. Lower oxygen saturation goals	P	P	P	X	X	X	X	X	X	X
13. High-saturation alarm set at 95%	P	P	P	X	X	X	X	X	X	X
14. Avoidance of routine suctioning for patients undergoing ventilation	P			P	X				X	P
15. Avoidance of hand-bagging for patients undergoing ventilation	P			P	X			X	X	
16. Nonroutine use of analgesics/sedatives for patients undergoing ventilation	P	P	P	P				X	P	P
17. Prophylactic use of methylxanthines before extubation				P			P	P	P	
18. Consensus regarding ventilatory management	P	P		P			P			
Nutrition/fluids										
19. Limited intravenous fluids	P	P	P		X	P		X	X	X
20. High-humidity environments	P	P		P	X	P				P
21. Limited volume expansion to treat low blood pressure	P	P		P	P	P	P	P		
22. Aggressive approach to patent ductus arteriosus	P	P		P				P	P	P
23. Early introduction of parenteral protein intake	P	P		P	P			P	P	P
24. Early introduction of lipids	P	P			P				P	P
25. Full total parenteral nutrition with increasing enteral feeding	P	P						P	P	P
26. Frequent use of human milk		P	P	P			P	P	P	
27. Vitamin A prophylaxis			P		P	X	P			P

P indicates a center already practicing this standard; X, an intervention selected by the center.

plementation. Specific practices at the benchmark centers and those selected by the intervention centers are shown in Table 1. More-detailed information on the potentially better practices, levels of evidence, and metrics used is contained in the Appendix.

Implementation of all selected potentially better practices was tracked with statistical control charts. Each intervention was assigned a predefined method of objective measurement based on observation of practices at the better-performing centers. For example, implementation of the use of high-saturation alarms was tracked with random audits of the use of alarms at the intervention sites and the control sites. The metrics used to track practice changes are summarized in the Appendix. Control charts (example shown in Fig 2) were generated by the data center with SAS software (SAS Institute, Cary, NC), provided to the intervention teams at 4- to 6-week intervals throughout the 2-year intervention period, and shared with all members of the NICUs to reinforce practice changes. Each team received control charts of all practices selected for implementation by any team. In this study, successful implementation of an intervention

FIGURE 2
Example of a control chart provided to intervention teams every 4 to 6 weeks to track progress in the implementation of potentially better practices. In this control chart, the practice of permissive hypercapnia is tracked with the metric mean $PaCO_2$ for intubated infants on day 3 of life. The horizontal line indicates the center's performance in the preintervention period; X, mean value for patients enrolled during the interval; circles, cumulative intervention mean; error bars, 95% CIs for the point estimates.
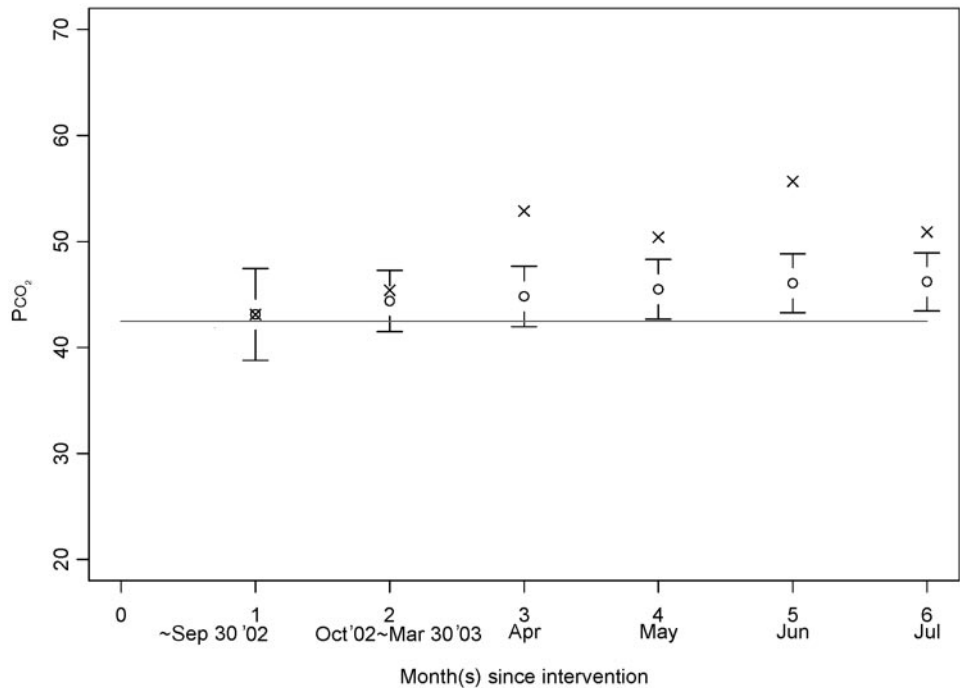
was defined as a statistically significant improvement from preintervention performance achieved within the 2-year intervention period and maintained at the end of the period. Data on practices in the best-performing centers before the study were available to the teams, but data on performance in the control centers were masked. Control centers were provided with annual summary data provided routinely to all network centers and were masked with respect to the work at the intervention centers. Control centers were prohibited from participating in other QI collaborative efforts focused on BPD. Because of ethical concerns with prohibiting practice changes in control centers during the 3-year trial period, local quality efforts initiated by clinicians who were not members of the research team were permitted. One control center changed its management approach to deemphasize mechanical ventilation and to emphasize nasal CPAP during the trial.

Outcomes

The primary outcome measure was the change in survival free of BPD between year 1 and year 3. BPD was assessed at postmenstrual age (PMA) of 36 weeks by using a validated physiologic definition that combined respiratory support and oxygen saturation and was developed for this trial.[13] Infants who were discharged before 36 weeks were assigned the diagnosis of BPD if discharged from the hospital with oxygen. For infants who were transferred before 36 weeks, a hierarchy was used to determine the diagnosis of BPD. If possible, a room air challenge was performed and the infant was assigned the diagnosis of BPD on the basis of the results

of the challenge. If a challenge was not possible, then the receiving institution was contacted and the infant was assigned the outcome of BPD if he or she was receiving oxygen supplementation, CPAP, or ventilation. If no information about the status at 36 weeks was available, then the infant was assigned the diagnosis of BPD if he or she was receiving oxygen supplementation, CPAP, or ventilation at the time of transfer. Overall, the outcome of BPD was determined for 99.9% of infants at intervention sites and 98.8% of infants at control sites. Secondary outcomes included death before hospital discharge, BPD severity (assessed with a modification of the National Institutes of Health consensus definition of BPD that included the physiologic definition), durations of mechanical ventilation, CPAP, and oxygen use, and length of hospital stay.[13,14] Other measures of common neonatal comorbidities were specified before the trial began and included severe intraventricular hemorrhage (Papile stage III or IV), cystic periventricular leukomalacia, severe retinopathy of prematurity (stage 3 or more), pneumothorax, patent ductus arteriosus, necrotizing enterocolitis (stage 2 or more), and late-onset sepsis (positive blood culture at >72 hours of age).[15–17] Arterial oxygen values, together with complete blood gas data and respiratory support information, were measured every 6 hours on days 1 to 7, at the values closest to 6 AM, noon, 6 PM, and midnight. The information was also recorded on days 14, 21, and 28 of life. All values were averaged. Severity of illness was assessed at 24 hours of age by using the Score of Neonatal Acute Physiology II.[18] Neonatal research nurses abstracted all data by using standardized definitions. Data were entered remotely

through electronic submission. Quality control procedures included range checking, internal comparisons for logic violations, and comparison of expected and observed values.

Human Subject Protection

The institutional review board at every site approved the study. One center provided families with a letter of information, and all others were given a waiver of consent requirements to collect deidentified data. The trial was registered at inception with the US National Library of Medicine trial registry (trial registration NCT00067613 [see www.clinicaltrials.gov]). In April 2003, data were reviewed by an independent data monitoring and safety committee, which recommended trial continuation.

Study Time Line and Statistical Methods

The preintervention year (study year 1) began in March 2001. Centers were assigned randomly to intervention or control in June 2001, to permit centers to free investigators for site visits in October and November 2001. Centers selected interventions, began implementation in May 2002, and continued interventions through a 2-year intervention period (study year 2 and year 3). Outcomes were compared between year 1 (March 2001 to May 2002) and year 3 (April 2003 to May 2004).

Analyses included all neonates with birth weights of <1250 g who were born at the centers and were free of major malformations. All analyses were based on an intent-to-treat model according to center assignment. Comparisons of the intervention and control centers were assessed by using mixed-model methods (SAS 9.1 loadable PROC GLIMMIX for binary outcomes and PROC MIXED for continuous outcomes), with the center entered as a random effect. These analyses accounted for the intraclass correlation within each center attributable to clustering from randomization according to center. The model for the analysis included the following terms: group (intervention or control), study year (year 1 or year 3), and group-year interaction. The group-year interaction term measures the difference between the 2 groups in changes from year 1 to year 3, which is the parameter of interest. A 0 coefficient for this term indicates a null treatment effect between the 2 groups. Other characteristics of the infants present at the time of birth were added to the model, including birth weight, gestational age of <26 weeks, race, gender, and prenatal steroid exposure. An interaction between gestational age (<26 weeks versus ≥26 weeks) and the main effect (group-year interaction) was added to the model if it showed a significant effect. Summary outcomes are shown when the gestational age interaction was not significant, and outcomes according to gestational age group are shown when the interaction was significant. Binary outcomes are presented as adjusted odds ratios

(ORs) for year 3 versus year 1, with 95% confidence intervals (CIs), and continuous outcomes as the adjusted difference of year 3 versus year 1.[19] A model with a term for severity of illness that included the Score for Neonatal Acute Physiology II with perinatal extension was also studied; the results obtained by using the additional term for severity of illness were not different from the results of the first model and are not shown.[18] We prespecified secondary analyses that evaluated the impact of the intervention according to center and according to gestational age (26 weeks versus ≥26 weeks). All study analyses were completed with SAS 9.1 (SAS Institute).

Sample Size

In this trial, with the center rather than the individual patient as the unit of randomization, sample size calculations accounted for both interhospital and intrahospital variability. The comparison of interest was the intervention/control difference in the change in rates of survival free of BPD between study years 1 and 3. The methods of Gail et al[20] from the Community Intervention Trial for Smoking Cessation were used in these calculations. Based on network data for 1999 and 2000, the rate of survival free of BPD (defined as oxygen use at 36 weeks) in inborn neonates of <1250 g was 45%. We calculated a sample size of 1400 neonates in each year of the study (100 patients per center in 14 centers) to yield 80% power (2-sided $\alpha = .05$) to detect an absolute difference of 14% in the change in rates of survival free of BPD between year 1 and year 3 in the intervention versus control groups. The magnitude of the effect selected was relatively large and was based on effect size seen in the only published work of collaborative QI to reduce BPD.[21]

RESULTS

Patient Population and Care Practices Before Intervention

The population for these analyses included 4095 live-born neonates with birth weights of 401 to 1250 g who were born at the 14 randomized centers between March 1, 2001, and April 30, 2004. Neonates who were not born at the centers and those with major malformations were excluded. Two neonates with major malformations were enrolled incorrectly and were removed subsequently, leaving 4093 neonates in the cohort. Among the 4093 neonates, 2871 were from year 1 and year 3 of the study and therefore were included in the analysis. The hospitals in the study arms were all level IIIB units with large volumes and accredited residency and neonatal/perinatal programs. Centers randomized to the control group were larger than those randomized to the intervention group, leading to more infants in the control centers (Fig 1). Infants in the intervention centers were slightly larger, were more mature, and included fewer white infants, compared with those in the control

centers (Table 2). Infants in the intervention centers were born to mothers who had received less prenatal steroid treatment (78% vs 87%; $P < .0001$), and the infants had higher severity of illness scores, as measured with the Score of Neonatal Acute Physiology II (23.8 ± 15.6 vs 20.9 ± 15.4; $P = .0002$). Despite these differences in infant characteristics, the incidence rates of BPD at 36 weeks, measured with the physiologic definition, were similar for the intervention and control centers (25.7% vs 28.3%; $P = .30$), as were the incidence rates of oxygen use at 36 weeks (38.5% vs 36.1%; $P = .40$).

Characteristics of the infants within centers did not differ across the 3 years of the trial with respect to birth weight, gestational age, gestational age of <24 weeks, gender, or prenatal steroid exposure. The only attribute with a statistically significant change was an increase in the percentage of black infants born at centers randomized to the intervention group from 30.6% to 36.3% ($P = .047$).

Interventions

The majority of intervention centers implemented their selected practices, although the rate of success did vary according to center (median rate of success: 75%; range: 40%–100%). The intervention group did change respiratory care practices (Table 3). Both intervention and control centers decreased the time of delivery of the first surfactant dose, with intervention centers decreasing from a median of 51 minutes to 31 minutes and control centers decreasing from a median of 41 minutes to 21 minutes. The intervention group significantly increased the use of CPAP on the first day of life (year 1: 16.9%; year 3: 24.2%) but, despite the 7.3% increase, usage was still below the rate of CPAP use in the control centers in both year 1 (26.5%) and year 3 (28.2%). Intervention centers also decreased the duration of mechanical ventilation in the first week of life (from 4.0 ± 2.7 days to 3.5 ± 2.8 days), whereas the control centers did not

change significantly (3.5 ± 2.8 days versus 3.4 ± 2.8 days). Despite having persistently higher rates of intubation on day 1 of life than did control centers between study year 1 and study year 3, intervention centers decreased the total duration of respiratory support by 5.3 days, whereas the control group decreased the duration by 4.1 days.

Intervention centers also implemented policies to reduce target oxygen saturations more frequently than did control centers. Five (71%) of 7 intervention centers and 2 (28.5%) of 7 control centers implemented use of high-saturation alarms to reduce oxygen saturation exposure. In monthly audits at all centers, patients receiving oxygen at intervention centers were more likely to have a high-saturation alarm in use (68.7% of audits with alarm in use; range: 39.8%–100%) than were those at control centers (10.5%; range: 0%–33%). This contributed to a reduction in the arterial oxygen levels measured in the first week of life in the intervention centers, from 74.1 ± 33.6 mm Hg (mean ± SD) in year 1 to 62.7 ± 21.2 mm Hg in year 3. In contrast, the values for the control centers were similar in years 1 and 3 (62.2 ± 30 vs 64.3 ± 26.4 mm Hg). Rates of postnatal steroid use declined significantly in both groups (intervention: from 14.3% to 4.4%; $P < .01$; control: from 14.3% to 5.6%; $P < .01$).

The final practice changes selected by intervention centers were restrictions of intravenous fluid volumes. Four centers selected this intervention, and 3 were successful in implementation, with reductions in delivered intravenous fluid (intervention: 126.8 mL/kg per day on day 3 of life in year 1 and 117 mL/kg per day in year 3; control: 125.1–122.6 mL/kg per day).

Outcomes

Intervention centers did not change the frequency of survival free of BPD faster than control centers (change score for intervention group adjusted for clustering:

TABLE 2 **Description of Populations at Intervention and Control Centers**

Characteristic	Intervention Centers ($N = 1752$)		Control Centers ($N = 2341$)	
	Year 1 ($N = 625$)	Year 3 ($N = 595$)	Year 1 ($N = 787$)	Year 3 ($N = 864$)
Birth weight, mean ± SD, g	932 ± 207	922 ± 206	899 ± 209	912 ± 212
PMA, mean ± SD, wk	27.3 ± 2.3	27.3 ± 2.5	27.0 ± 2.3	27.1 ± 2.5
PMA of <26 wk, n (%)	151 (24.2)	157 (26.4)	229 (28.8)	244 (28.2)
Male gender, n (%)	338 (54.1)	311 (52.3)	394 (49.9)	432 (50.2)
Race, n (%)				
White	196 (31.4)[a]	170 (28.6)[a]	329 (41.8)	365 (42.3)
Black	191 (30.6)	216 (36.3)	368 (46.8)	377 (43.6)
Other	238 (38.1)	209 (35.1)	90 (11.4)	122 (14.2)
Maternal age, mean ± SD, y	27.2 ± 6.4	27.3 ± 6.7	27.2 ± 6.4	27.0 ± 6.3
Prenatal steroid exposure, n (%)	482 (77.1)	470 (79.1)	680 (86.4)	761 (88.2)
Score of Neonatal Acute Physiology II, mean ± SD	23.4 ± 15.0	24.6 ± 16.8	22.4 ± 14.6	19.5 ± 15.7

Comparisons of population characteristics in each group between study year 1 and study 3 were not significantly different, except for race in the intervention centers.

[a] $P = .047$.

−0.01; 95% CI: −0.06 to 0.04; control group: +0.003; 95% CI: −0.04 to 0.05). The changes between study year 1 and study year 3 in overall survival rates, rates of survival free of BPD, and severity of BPD did not differ between the intervention and control centers (Table 4 and Fig 3). Infants cared for in intervention centers did not differ in the incidence of common neonatal morbidities, compared with those cared for in control centers (Tables 3 and 5). There was a nonsignificant trend toward increased rates of severe intraventricular hemorrhage in the intervention centers between year 1 and year 3 (14.4% vs 18.1%), compared with the control centers (13.3% vs 14.1%). The incidence rates of periventricular leukomalacia, severe retinopathy of prematurity, necrotizing enterocolitis, patent ductus arteriosus, and growth failure were similar between the 2 groups in study years 1 and 3.

Effects of Intervention According to Gestational Age

During study design, we hypothesized that the intervention would be less successful for less-mature infants because of increased biological vulnerability. Therefore, we analyzed outcomes for 2 gestational age groups, that is, PMA of <26 weeks and PMA of ≥26 weeks. As anticipated, for infants with PMA of ≥26 weeks across the 3 years of the trial, rates of survival to PMA of 36 weeks (94% vs 66%) and survival free of BPD (74% vs 30%) were significantly higher than those for infants with PMA of <26 weeks. Overall, results did differ according to PMA group (Fig 4). For infants with PMA of ≥26 weeks, there were no differences in rates of survival to PMA of 36 weeks between intervention and control centers, but a statistically nonsignificant trend toward reduced BPD rates was seen in intervention centers. In contrast, in the subgroup of infants with PMA of <26 weeks,

trends toward reduced survival rates and increased rates of BPD were seen at the intervention centers (Table 4 and Fig 4). Of note, this difference was driven by a lower-than-anticipated rate of BPD in year 1. The interaction between PMA and treatment assignment was significant (P = .01) (gestational age of <26 weeks: BPD in intervention: OR: 2.53; 95% CI: 1.31–4.90; BPD in control: OR: 1.01; 95% CI: 0.63–1.63; gestational age of ≥26 weeks: BPD in intervention: OR: 0.80; 95% CI: 0.56–1.16; BPD in control: OR: 0.96; 95% CI: 0.70–1.33).

Outcomes According to Center

Results did differ according to center in both the intervention and control groups. In the intervention group, 1 center had a significantly improved rate of survival free of BPD (OR: 1.93; 95% CI: 1.00–3.74) and 6 centers had no significant change (OR: 0.37–1.22). However, in 2 of the centers with no significant change overall, there was a significant interaction between the intervention and PMA. In those 2 centers, no change in BPD rates was seen for neonates born at PMA of ≥26 weeks (OR: 1.33; 95% CI: 0.75–2.36) but worsened outcomes were seen for infants born at PMA of <26 weeks (OR: 0.18; 95% CI: 0.06–0.55). Those 2 centers drove the gestational age differences described in the overall trial. In the control centers, 1 center significantly improved the rate of survival free of BPD (OR: 1.96; 95% CI: 1.00–3.84), whereas 5 centers showed no significant change (OR: 0.40–1.35) and 1 center worsened significantly (OR: 0.53; 95% CI: 0.29–0.99).

DISCUSSION

We showed that, in a rigorous trial using the center as the unit of randomization, centers that implemented QI processes were successful in changing care practices but did not improve rates of survival free of BPD for high-

TABLE 3 Respiratory Outcomes for Intervention and Control Groups

Variables	Benchmark Centers, Year 1	Intervention Centers			Control Centers			P
		Year 1	Year 3	Change Score/ OR (95% CI), Year 3 vs Year 1[a]	Year 1	Year 3	Change Score/ OR (95% CI), Year 3 vs Year 1[a]	
Continuous								
Total duration of ventilation, mean ± SD, d	13.8 ± 19.7	15.0 ± 18.2	14.3 ± 19.3	−1.2 (−3.1 to 0.7)	16.1 ± 21.3	14.9 ± 19.5	−0.9 (−2.5 to 0.7)	.79
Duration of ventilation in first 7 d, mean ± SD, d	3.5 ± 2.8	4.0 ± 2.7	3.5 ± 2.8	−0.6 (−0.8 to −0.3)	3.5 ± 2.8	3.4 ± 2.8	−0.0 (−0.3 to 0.2)	.0026
Duration of CPAP, mean ± SD, d	6.6 ± 9.5	8.5 ± 11.9	8.0 ± 11.4		9.4 ± 11.2	9.3 ± 11.5		.21
Duration of oxygen therapy, mean ± SD, d	28.1 ± 26.7	33.2 ± 27.0	29.0 ± 27.0	−4.9 (−7.6 to −2.2)	34.3 ± 27.9	31.7 ± 27.7	−2.2 (−4.5 to 0.1)	.14
Dichotomous								
CPAP use on day 1 of life, %	15.8	16.9	24.2	1.71 (1.27–2.30)	26.5	28.2	1.09 (0.85–1.39)	.08
Surfactant use, %	66.6	75.8	76.9	0.96 (0.71–1.30)	73.5	74.9	1.16 (0.90–1.49)	.35
Pneumothorax, %	6.7	8.1	7.4	0.85 (0.55–1.31)	5.7	6.1	1.10 (0.72–1.68)	.40

See "Methods" for elements included in the model

[a] For continuous variables, the model compared the differences in the change scores between year 1 and year 3 for intervention and control groups; for dichotomous variables, the model compared the differences in the ORs.
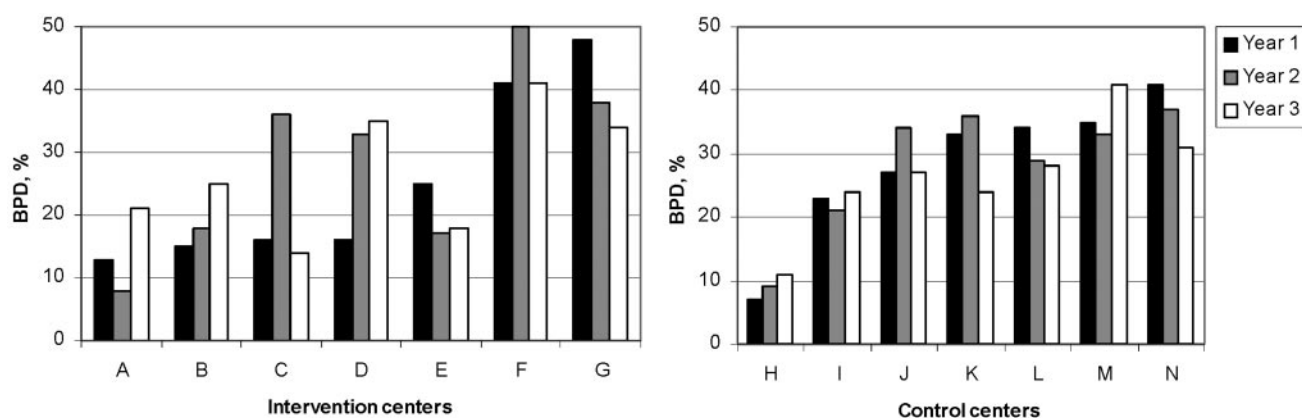
FIGURE 3
Change in rates of BPD according to study year in the intervention and control centers. Each center is shown in each of the 3 study years.

TABLE 4 Primary Outcomes for Benchmark, Intervention, and Control Groups

Outcome	Benchmark Centers, Year 1	Intervention Centers				Control Centers				P
		%		OR (95% CI)		%		OR (95% CI)		
		Year 1	Year 3			Year 1	Year 3			
Survival free of BPD[a]	73.3	63.3	62.2			62.7	62.8			.40 (.03)[b]
Gestational age of <26 wk	52.8	38.4	22.3	0.47 (0.27–0.82)		31.7	29.5	0.84 (0.55–1.30)		.10
Gestational age of ≥26 wk	83.1	71.3	76.5	1.23 (0.88–1.71)		75.3	75.9	1.00 (0.75–1.33)		.35
Survival to PMA of 36 wk	84.4	85.3	84.2	0.96 (0.67–1.37)		87.5	86.3	0.83 (0.60–1.14)		.56
BPD, physiologic definition[a]	13.4	25.7	26.1			28.3	27.2			.13 (.01)[b]
Gestational age of <26 wk	24.0	37.6	60.7	2.53 (1.31–4.90)		55.3	55.8	1.01 (0.63–1.63)		.03
Gestational age of ≥26 wk	9.5	23.1	18.7	0.80 (0.56–1.16)		20.1	19.2	0.96 (0.70–1.33)		.47
BPD severity[a]				0.81 (0.60–1.11)				0.98 (0.76–1.28)		.36
Severe	9.8	11.3	12.2			14.8	13.0			
Moderate	3.3	14.3	13.8			12.8	13.9			
Mild	0.8	8.1	4.4			4.5	4.7			
None	86.2	66.2	69.5			67.0	67.9			
Undefined	0	0	0.1			0.8	0.4			

Benchmark center data were collected from October 2000 to May 2001. All analyses compared the differences in the changes between study year 1 and year 3 in the intervention centers, compared with the control centers. Variables included in the model were group assignment (intervention versus control), study year, group-study year, birth weight, gender, race, prenatal steroid exposure, interaction with gestational age of <26 weeks, and random clustering effect according to center.

[a] BPD severity was defined with a modification of the National Institutes of Health consensus definition of BPD that included results of a timed room air challenge (see "Methods"). For this analysis, BPD severity was dichotomized as any BPD versus no BPD.

[b] Analyses that demonstrated effect modification by gestational age are shown with overall results and with results according to gestational age group. The significance of interaction term is shown in parentheses.

risk preterm neonates, compared with control centers. Intervention centers benefited from working together over time, tracking their results, and changing practices. Intervention centers modified their practices to be more similar to those of the benchmark centers. Intervention centers significantly decreased measured oxygen tension in the first week of life and significantly increased the use of nasal CPAP, compared with control centers. Despite these changes in delivery room and respiratory care practices, we did not demonstrate increased rates of survival free of BPD. In part, unexpectedly low rates of BPD in the preintervention year in the intervention centers made improvements difficult.

By chance, centers randomized to intervention had higher rates of endotracheal intubation (77% vs 66%),

lower rates of CPAP on day 1 of life (16.9% vs 26.5%), and more days of mechanical ventilation in the first 7 days of life (4.0 ± 2.7 vs 3.5 ± 2.8 days). The goal of the study was to assess the utility of benchmarking to accelerate practice changes. As is normal in QI studies, intervention centers were focused initially on the performance of the 3 benchmark centers and later on performance at their own centers. Intervention centers were masked with respect to processes at the control centers. Differences in practices at centers were anticipated in the design and represent the reason why a change score between year 1 and year 3 was used as the primary outcome statistic; such an approach corrects for potential baseline differences in patient characteristics and practices. Despite differences in practices, interven-

TABLE 5 Other Neonatal Outcomes

	Intervention Centers			Control Centers			Adjusted P
	%		OR (95% CI), Year 3 vs Year 1	%		OR (95% CI), Year 3 vs Year 1	
	Year 1	Year 3		Year 1	Year 3		
Severe intraventricular hemorrhage	14.4	18.1		13.3	14.1		0.28 (.03)[a]
PMA of <26 wk	26.2	45.6	2.33 (1.38–3.94)	26.0	28.6	1.12 (0.72–1.75)	.04
PMA of ≥26 wk	10.8	8.8	0.83 (0.53–1.31)	8.4	8.7	1.05 (0.68–1.60)	.47
Periventricular leukomalacia	1.5	1.9	1.23 (0.50–3.00)	2.2	1.5	0.68 (0.32–1.47)	.33
Retinopathy of prematurity, stage ≥3	13.3	16.4	1.49 (1.01–2.26)	16.3	17.0	1.10 (0.79–1.54)	.26
Necrotizing enterocolitis, stage ≥2	7.8	8.9	1.10 (0.73–1.66)	9.1	10.3	1.19 (0.85–1.66)	.77
One or more of above morbidities	84.6	84.3	0.96 (0.68–1.35)	84.8	83.7	0.97 (0.72–1.32)	.95
Discharged with oxygen	19.6	12.6		8.4	9.2		.34 (.02)[a]
Gestational age of <26 wk	25.8	28.6	0.88 (0.32–2.42)	24.8	30.2	1.28 (0.65–2.52)	.55
Gestational age of ≥26 wk	15.6	10.6	0.68 (0.43–1.08)	5.1	5.1	0.94 (0.52–1.68)	.41

For dichotomous variables, the model compared the differences in the ORs for year 3 versus year 1 for the intervention and control groups. See "Methods" for elements included in the model.

[a] Variable shows significant effect modification by gestational age ($P = .03$); overall results and results according to gestational age group are shown. The significance of interaction term is shown in parentheses.
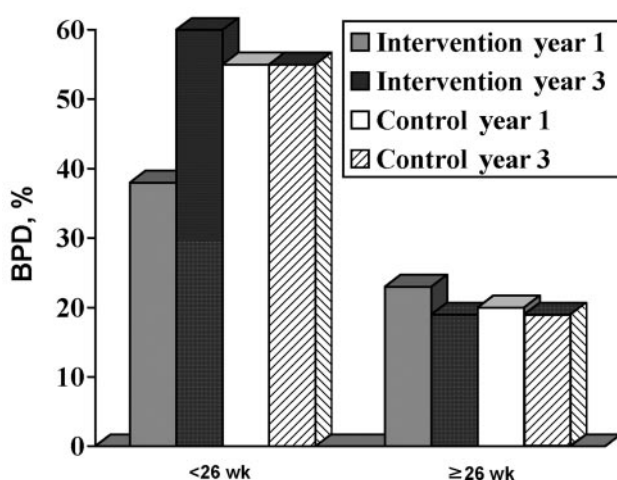
FIGURE 4

Impact of PMA on rates of BPD in study years 1 and 3 in the intervention and control centers. Intervention centers demonstrated a low rate of BPD in the preintervention year that increased in intervention year 3 to a level equivalent to that in control sites.

tion centers did not improve rates of survival free of BPD faster than control centers. If in fact the practices on which the centers focused were important determinants of BPD, then the intervention centers should have been biased toward a greater likelihood of finding an improvement, because they had greater opportunities to improve practice.

BPD is a complex disease with multifactorial pathophysiologic processes.[6,11,14] Disruption of the development of fragile lung parenchyma at a critical period of alveolar and vasculature maturation is thought to be a primary determinant of BPD. Key contributors to such disruption are oxidative injury from oxygen, ventilator-induced lung injury, and inflammation.[11,22–28] Interventions in this trial reduced ventilator pressures, reduced time with mechanical ventilation, and reduced oxygen concentrations. Interventions to avoid volutrauma were supported by evidence from preclinical studies in animals and human neonates and were reinforced by data and direct observation at the benchmark centers.[25–28] However, as shown in the Appendix, the strength of the evidence for these interventions in human neonates is weak, with few randomized trials. Existing evidence equally supports 2 philosophically divergent interventions, that is, endotracheal intubation and the early delivery of surfactant versus avoidance of endotracheal intubation and the use of CPAP.[26,27] It is currently unclear which is the superior approach. Concern regarding potential injury from oxidative stress from high-oxygen environments for which preterm neonates are poorly prepared biologically led centers to emphasize oxygen reduction, the only intervention selected by all 7 intervention sites. In a posthoc analysis using both intervention and control centers, interventions focused on pressure reduction rather than oxygen reduction were more successful in reducing BPD rates (data not shown).

Although interest in QI to improve health care and outcomes is not new, rigorous randomized trials evaluating the method have been conducted only recently, with mixed results. The majority of the trials focused on improving delivery of evidence-based services to adults (eg, use of β-receptor blockers) or improving the efficiency of care delivery (eg, reducing waiting times). Several trials tested the utility of multimodal interventions. In a cluster-randomized trial, Ferguson et al[29] demonstrated improved preoperative β-receptor blocker and internal mammary graft use in a nationwide QI effort. Although a statistically significant increase in prescription of preoperative β-receptor blocker therapy was seen, the magnitude of the increase was modest (7.3% vs 3.6% in control centers). Mehta et al[30] tested a multimodal intervention led by local opinion leaders to measure the impact on 11 indicators of the quality of acute myocardial infarction care. Some indicators improved in intervention centers, whereas others improved more in the control centers. Overall, the absolute gains ranged

from 4% to 12%. Kiefe et al[31] used "achievable benchmarks of care," that is, levels of performance achieved by top-performing centers. Those receiving the benchmark feedback improved delivery of influenza vaccine by 18%, compared with the control group. In contrast to these positive results, other trials seeking to improve delivery of β-receptor blocker use after myocardial infarction, clinical preventive services, compliance with national guidelines for the treatment of hypertension and depression, and compliance with protocols for the care of patients with AIDS failed to demonstrate significant changes in practice.[32–35] A recently published meta-analysis of QI strategies for patients with diabetes mellitus demonstrated that most trials generated only modest improvements in glycemic control.[4] The investigators also found evidence strongly suggesting publication bias, with smaller studies being more likely to show positive effects than larger studies.

Trials of QI in neonatology and pediatrics are more limited. Lozano et al[36] reported the results of an intensive QI intervention conducted by Pediatric Asthma Care Patient Outcomes Research Team II. A resource-intensive intervention using organizational change plus physician peer leader education was more effective than physician education alone.[36] In neonatology, the Vermont Oxford Network used multifaceted QI techniques to improve patient outcomes focused on rates of nosocomial infections (5.5% reduction, compared with 1.6% for nonparticipants) and BPD (12.5% reduction from 43.5% to 31%, compared with 8.3% reduction for nonparticipants).[21] The Vermont Oxford Network group reported a subsequent study that enrolled self-selected centers focused on BPD and showed reductions in rates of BPD in before/after comparisons.[37] In another study, the Vermont Oxford Network investigators demonstrated outcomes similar to those of the current trial.[38] Implementation of a multimodal QI intervention resulted in earlier administration of surfactant, compared with control centers, but those improved practices did not translate into improved patient outcomes, measured as death or pneumothorax.

One important difference between the current trial and previous studies of QI to reduce BPD rates was the outcome measure we used. Previous studies used a clinical definition of BPD defined by oxygen and/or ventilation exposure at 36 weeks, without controlling for oxygen saturation values delivered. As a prelude to this trial, we developed a rigorous definition of BPD that included a room air challenge for selected infants (those receiving <30% effective oxygen).[13] This physiologic definition of BPD was applied equally in the intervention and control centers. As we reported previously, the definition resulted in a mean reduction of 10% in the rates of BPD (range: 0%–44%). The implementation of this definition could be considered an intervention that focused clinicians' attention on the importance of integrating oxygen delivery, especially through a nasal cannula, with saturation monitoring at both intervention and control centers.[39] It might have been the most effective intervention in the trial, dwarfing the effects of other potentially beneficial practices.

What accounts for our finding that the multimodal intervention failed to improve patient outcomes? One possibility is that, despite randomization, there were important random differences between intervention and control centers in the preexisting rates of survival free of BPD. However, the rates were comparable (63.3% in the intervention centers and 62.8% in the control centers in year 1). Another possibility is that our trial was not large enough to identify a clinically important benefit from benchmarking. The 95% CIs for the changes in rates of survival free of BPD resulting from benchmarking and QI excluded a benefit greater than a 4.4% improvement in the rate of survival without BPD and included a hazard as great as a 6.1% increase in the rate of death or BPD. These CIs indicate that the trial was large enough to exclude important larger effects.[40,41] Another possibility is that the QI training was ineffective. We think that this was not the case, because intervention centers demonstrated greater practice changes than did control centers.

A final (and we think more likely) possibility is that adopting practices from centers with exemplary outcomes may not be beneficial when there is only weak evidence supporting these practices. Well-controlled studies reporting benefits from benchmarking largely have been studies promoting the use of interventions established previously as beneficial in randomized trials. The interventions in our study were those with the strongest available evidence. However, few (such as early administration of surfactant) have been shown to be beneficial in randomized trials. Of the myriad of practice differences between centers, it remains to be established whether the practices that result in superior outcomes in benchmark centers can be reliably recognized and implemented by visiting health care teams. It is even possible that some interventions selected in QI efforts affect outcomes adversely.[42,43] An example of this is the selection of a skin emollient as an intervention to reduce infection by one Vermont Oxford Network collaborative group. Emollient was shown to reduce infection rates in a single-center study but was later shown to increase infection rates in a large randomized trial.[44,45] Introducing change, no matter how well intentioned, may perturb a stable system, with potentially adverse outcomes. The apparent increase in rates of severe intraventricular hemorrhage among neonates with PMA of <26 weeks in intervention centers may be a statistical anomaly or a real but unintended adverse consequence of changes in care. It is possible that interventions in the delivery room prolonged the time spent in the delivery

room and contributed to hypothermia and associated intraventricular hemorrhage.

In this cluster-randomized, controlled trial, NICU teams trained in benchmarking and QI techniques benefited from the intervention with practice changes but did not improve rates of survival free of BPD in neonates with birth weights of <1250 g, compared with centers continuing usual practice. These results have implications for the design of future QI trials, in that other interventions may be required to produce change. Additional refinements are needed to create and to maintain larger magnitudes of change and to improve patient outcomes.

APPENDIX: DESCRIPTION OF POTENTIALLY BETTER PRACTICES

Practices in the Delivery Room

1. Fellow or attending physician present at every high-risk delivery. Resuscitation of high-risk infants was led by a fellow or attending physician at the benchmark sites. Level of evidence: not available; metric: audit during site visits.

2. Respiratory therapist present at every high-risk delivery. The resuscitation team included a respiratory therapist at the benchmark sites. Level of evidence: not available; metric: audit during site visits.

3. Consistent equipment in all delivery rooms. Each resuscitation site was configured with identical equipment, to facilitate the resuscitation of high-risk patients. Level of evidence: not available; metric: audit during site visits.

4. Limited tidal volumes used in manual ventilation during resuscitation. At the benchmark sites, teams focused on limiting tidal volumes during resuscitation by assessing chest wall excursion visually or measuring delivered tidal volumes. The goal was to have barely visible chest wall movements. Evidence in animal models supports this concept. Level of evidence: level 2C; metric: audit during site visits; metric used: first peak inspiratory pressure on admission to the NICU.

5. Prophylactic use of surfactant. At 2 benchmark centers, infants at gestational ages of <28 weeks were immediately intubated and given surfactant. The third center emphasized CPAP beginning in the delivery room. Strong evidence supports this concept to minimize the severity of respiratory distress syndrome. Level of evidence: level 1A[46]; metric: time to surfactant use in the delivery room.

6. Use of device to provide positive end-expiratory pressure and to limit tidal volume. This practice was not in place at the benchmark centers. Intervention teams added this to the list of potentially better practices to support limitation of delivered tidal volumes. Level of evidence: not available; metric: audit during site visits.

Respiratory Care Practices

7. Selective intubation with liberal use of CPAP. One benchmark center emphasized CPAP beginning in the delivery room. The other 2 centers used intubation with prophylactic surfactant treatment. The evidence for a primary CPAP strategy is weak, with reports from small case series and observational studies. Level of evidence: indeterminate[47]; metric: proportion of infants treated with CPAP on admission to the NICU.

8. Early use of surfactant if intubated. The benchmark center using a primary CPAP strategy administered surfactant at once if a decision to intubate the infant was made. Evidence supports this concept to minimize the severity of respiratory distress syndrome. Level of evidence: level 1[48]; metric: time to surfactant administration in the NICU.

9. Assessment of volume/pressure and targeting of lowest levels to achieve modest chest rise and to avoid exuberant chest wall motion if intubated. All 3 benchmark centers focused on low tidal volumes for intubated infants. All 3 centers used pressure-controlled, time-cycled ventilators. None had the capacity to measure tidal volume, and instead they used physical examination and assessment of P_{CO_2} to limit tidal volumes. Level of evidence: level 5B; metric: mean peak inspiratory pressure on days 1 and 3 for all intubated neonates.

10. Aggressive weaning and early extubation if intubated. Two benchmark centers weaned patients aggressively and extubated them without birth weight or postnatal age limitations. In these 2 centers, teams noted during the site visits that it was common to see tiny infants at 24 hours of age receiving CPAP. The third center did not extubate patients but weaned them in the first 24 hours to low ventilatory rates and tidal volumes that were comparable to those delivered with CPAP. Level of evidence: level 5; metric: duration of ventilation in the first 7 days of life.

11. Higher Pa_{CO_2} targets for all patients. The 3 benchmark centers accepted higher Pa_{CO_2} levels to permit weaning from ventilators and/or the use of CPAP. Experimental animal data and uncontrolled human observational studies support permissive hypercapnia as a protective strategy. Level of evidence: indeterminate[49]; metric: mean Pa_{CO_2} for all patients with measurements on days 1 and 3 of life.

12. Lower oxygen saturation goals. The 3 benchmark centers accepted lower oxygen saturation targets of 85% to 90%. In addition, caregivers were noted, during the site visits, to observe infants during desaturation events without increasing oxygen supplementation, with the goal of allowing the infant to resolve the desaturation event independently. Level of evidence: indeterminate[50]; metric: mean Pa_{O_2} measured at 4 time points daily during the first 7 days of life for all those with measurements.

13. High-saturation alarm set at 95%. The 3 bench-

mark centers set the oxygen saturation alarm at 95% and rapidly weaned patients from supplemental oxygen when the saturation range exceeded the target. Level of evidence: not available; metric: monthly random audits of enabled saturation alarms for all infants enrolled in the trial who were receiving oxygen at control and intervention sites.

14. Avoidance of routine suctioning for patients undergoing ventilation. The 3 benchmark centers avoided suctioning for patients undergoing ventilation that was set by a time schedule and instead assessed the patients at intervals and suctioned as needed. In addition, 2 centers used an inline suction system. Level of evidence: indeterminate[51]; metric: audit during site visits.

15. Avoidance of hand-bagging for patients undergoing ventilation. One benchmark center prohibited the practice of ventilating with an anesthetic or self-inflating bag, as a method to limit exposure to unregulated tidal volumes. Level of evidence: not available; metric: audit during site visits.

16. Nonroutine use of analgesics/sedatives for patients undergoing ventilation. None of the benchmark centers routinely administered analgesics/sedatives to patients treated with mechanical ventilation. Instead, comfort techniques such as swaddling were used. Level of evidence: indeterminate[52]; metric: audit during site visits.

17. Prophylactic use of methylxanthines before extubation. Routine administration of methylxanthines before extubation was not used at the benchmark centers. Level of evidence: level 2; metric: not selected.

18. Consensus regarding ventilatory management. At 2 benchmark centers, there was high consistency in ventilator management practices among individual physicians and teams. Level of evidence: level 2[53]; metric: not selected.

Fluid and Nutrition Practices

19. Limited intravenous fluids. At all 3 benchmark centers, intravenous fluids were initiated at 80 to 100 mL/kg per day and were adjusted by using daily weight goals. The intent was for weight loss to occur in the first 7 days of life. Level of evidence: level 2[54]; metrics: mean intravenous fluid intake on days 1 and 3 and percentages of weight loss on days 3 and 7.

20. High-humidity environments. Two benchmark centers used high-humidity environments to limit intravenous fluid administration. Level of evidence: level 5; metric: audit during site visits.

21. Limited volume expansion to treat low blood pressure. One benchmark center used protocols to decrease treatment of low blood pressure, to limit intravenous fluid administration. Level of evidence: level 5; metric: not selected.

22. Aggressive approach to patent ductus arteriosus. One benchmark center used prophylactic indomethacin treatment, and 1 center screened for patent ductus arteriosus and treated patients with indomethacin in the first 24 hours. The third center had high rates of patent ductus arteriosus ligation. The patent ductus arteriosus was ligated for patients who experienced failure of indomethacin treatment and had persistent oxygen requirements exceeding 40%. Level of evidence: level 5; metric: not selected.

23. Early introduction of parenteral protein intake. Two benchmark centers began total parenteral nutrition administration on admission to the NICU. Level of evidence: level 5; metric: not selected.

24. Early introduction of lipids. Two benchmark centers began intravenous lipid administration by 24 hours of age. Level of evidence: indeterminate; metric: not selected.

25. Full total parenteral nutrition with increasing enteral feeding. Two benchmark centers maintained total parenteral nutrition at 100 to 120 mL/kg per day as enteral nutrition was increased. Level of evidence: not available; metric: not selected.

26. Frequent use of human milk. Two benchmark centers promoted the use of human milk and had programs in place to support human milk feedings. Both centers had human milk administered to >60% of their patients. Level of evidence: indeterminate[55]; metric: not selected.

27. Vitamin A prophylaxis. One benchmark center used vitamin A prophylaxis. The other 2 centers chose not to implement prophylaxis because their BPD rates were low and nurses objected to intramuscular injections. Level of evidence: level 2; metric: audit during site visits.

REFERENCES

1. Institute of Medicine, Committee on Quality of Health Care in America. *To Err Is Human: Building a Safer Health System.* Washington, DC: National Academies Press; 2000

2. Institute of Medicine, Committee on Quality of Health Care in America. *Crossing the Quality Chasm: A New Health System for the 21st Century.* Washington, DC: National Academies Press; 2001

3. Institute of Medicine, Committee on the Health Professions Education Summit. *Health Professions Education: A Bridge to Quality.* Washington, DC: National Academies Press; 2003

4. Shojania KV, Ranji SR, McDonald JM, et al. Effects of quality improvement strategies for type 2 diabetes on glycemic control. *JAMA.* 2006;296:427–440

5. Lemons JA, Oh W, Korones SB, et al. Very low birth weight outcomes of the National Institute of Child Health and Human Development Neonatal Research Network, January 1995 to December 1996. *Pediatrics.* 2001;107(1). Available at: www.pediatrics.org/cgi/content/full/107/1/e1

6. Kinsella JP, Greenough A, Abman SH. Bronchopulmonary dysplasia. *Lancet.* 2006;367:1421–1431

7. Vohr BR, Wright LL, Dusick AM, et al. Neurodevelopmental and functional outcomes of extremely low birth weight infants in the National Institute of Child Health and Human Development Neonatal Research Network, 1993–1994. *Pediatrics.* 2000;105:1216–1226

8. Furman L, Borowoski CE, Aucott S, Hack M. Hospitalization as a measure of morbidity among very low birth weight infants with chronic lung disease. *J Pediatr.* 1996;128:447–452

9. Kennedy KA, Verter J, Tyson JE, et al. Center differences in survival without chronic lung disease (CLD) in very-low-birth-weight (VLBW) infants are not explained by population differences, or the presence and early management of respiratory distress syndrome (RDS). *Pediatr Res.* 1997;40:256A

10. Vohr BR, Wright LL, Dusick AM, et al. Center differences and outcomes of extremely low birth weight infants. *Pediatrics.* 2004;113:781–789

11. Van Marter LJ Allred EN, Pagano M, et al. Do clinical markers of barotrauma and oxygen toxicity explain interhospital variation in rates of chronic lung disease? *Pediatrics.* 2001;105:1194–1201

12. Plesk P. Innovative thinking for the improvement of medical systems. *Ann Intern Med.* 1999;131:438–444

13. Walsh M, Yao Q, Gettner P, et al. Impact of a physiologic definition on bronchopulmonary dysplasia rates. *Pediatrics.* 2004;114:1305–1311

14. Jobe AH, Bancalari E. Bronchopulmonary dysplasia. *Am J Respir Crit Care Med.* 2001;163:1723–1729

15. Papile LA, Burstein J, Burstein R, Koffler H. Incidence and evolution of subependymal and intraventricular hemorrhage: a study of infants with birth weights less than 1,500 gm. *J Pediatr.* 1978;92:529–534

16. Committee for the Classification of Retinopathy of Prematurity. An international classification of retinopathy of prematurity. *Arch Ophthalmol.* 1984;102:1130–1134

17. Walsh MC, Kliegman R. Necrotizing enterocolitis: treatment based on staging criteria. *Pediatr Clin North Am.* 1986;33:179–201

18. Richardson DK, Corcoran JD, Escobar GJ, Lee SK. SNAP-II and SNAPPE-II: simplified newborn illness severity and mortality risk scores. *J Pediatr.* 2001;138:92–100

19. Zeger SL, Liang KY. An overview of methods for the analysis of longitudinal data. *Stat Med.* 1992;11:1825–1839

20. Gail MH, Byar DP, Pechacek TF, Corle DK. Aspects of statistical design for the Community Intervention Trial for Smoking Cessation (COMMIT). *Control Clin Trials.* 1992;13:6–21

21. Horbar JD, Rogowski J, Plsek PE, et al. Collaborative quality improvement for neonatal intensive care. *Pediatrics.* 2001;107:14–22

22. Jobe AH, Ikegami IM. Antenatal infection/inflammation and postnatal lung maturation and injury. *Respir Res.* 2001;2:27–32

23. Albertine KH, Jones GP, Starcher BC, et al. Chronic lung injury in preterm lambs: disordered respiratory tract development. *Am J Respir Crit Care Med.* 1999;159:945–958

24. Askie LM, Henderson-Smart DJ, Irwig L, Simpson JM. Oxygen-saturation targets and outcomes in extremely preterm infants. *N Engl J Med.* 2003;349:959–967

25. Tremblay LV, Velena F, Ribeiro S, et al. Injurious ventilatory strategies increase cytokines and c-*fos* mRNA expression in an isolated rat lung model. *J Clin Invest.* 1997;99:944–952

26. Linder W, Vosbeck S, Hummler H, Pohlandt F. Delivery room

management of extremely low birth weight infants: spontaneous breathing or intubation. *Pediatrics.* 1999;103:961–967

27. Aly HZ. Nasal prongs continuous positive airway pressure: a simple yet powerful tool. *Pediatrics.* 2001;108:759–761

28. Ikegami M, Jobe AH. Postnatal lung inflammation increased by ventilation of preterm lambs exposed antenatally to *Escherichia coli* endotoxin. *Pediatr Res.* 2002;52:356–362

29. Ferguson TB Jr, Peterson PE, Coombs LP, et al. Use of continuous quality improvement to increase use of process measures in patients undergoing coronary artery bypass graft surgery. *JAMA.* 2003;290:49–56

30. Mehta RH, Montoye CK, Gallogly M, et al. Improving quality of care for acute myocardial infarction: the Guidelines Applied in Practice (GAP) Initiative. *JAMA.* 2002;287:1269–1276

31. Kiefe CI, Allison AJ, Williams OD, et al. Improving quality improvement using achievable benchmarks for physician feedback: a randomized controlled trial. *JAMA.* 2001;285:2871–2879

32. Bradley EH, Herrin HJ, Mattera JA, et al. Quality improvement efforts and hospital performance: rates of β-blocker prescription after acute myocardial infarction. *Med Care.* 2005;43:282–292

33. Solberg LI, Kottke T, Brekke ML, et al. Failure of a continuous quality improvement intervention to increase the delivery of preventive services: a randomized trial. *Eff Clin Pract.* 2000;3:105–115

34. Goldberg HI, Wagner EH, Fihn SD, et al. A randomized controlled trial of CQI teams and academic detailing: can they alter compliance with guidelines? *Jt Comm J Qual Improv.* 1998;24:130–142

35. Landon BE, Wilson IB, McInnes K, et al. Effects of a quality improvement collaborative on the outcome of care of patients with HIV infection: the EQHIV Study. *Ann Intern Med.* 2004;140:887–896

36. Lozano P, Finkelstein JA, Carey VJ, et al. A multisite randomized trial of the effects of physician education and organizational change in chronic-asthma care: health outcomes of the Pediatric Asthma Care Patient Outcomes Research Team II Study. *Arch Pediatr Adolesc Med.* 2004;158:875–883

37. Payne NR, Finkelstein MJ, Chen S, et al. Reduction of chronic lung disease (CLD) among very low birth weight (VLBW) infants: experience of a Vermont Oxford Network (VON) quality improvement collaborative (NIC/Q 2002). *E-PAS.* 2005;57:134

38. Horbar JD, Carpenter J, Buzas J, et al. Collaborative quality improvement to promote evidence based surfactant for preterm infants: a cluster randomised trial. *BMJ.* 2004;329:1004–1010

39. Walsh M, Engle W, Laptook A, et al. Oxygen delivery through nasal cannulae to preterm infants: can practice be improved? *Pediatrics.* 2005;116:857–861

40. Goodman SN, Berlin JA. The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. *Ann Intern Med.* 1994;121:200–206

41. Detsky AS, Sackett DL. When was a "negative" clinical trial big enough? How many patients you needed depends on what you found. *Arch Intern Med.* 1985;145:709–712

42. Wachter RM. Expected and unanticipated consequences of the quality and information technology revolutions. *JAMA.* 2006;295:2780–2783

43. Han YY, Carcill JA, Venkatarman ST, et al. Unexpected increased mortality after implementation of a commercially sold computerized physician order entry system. *Pediatrics.* 2005;144:1506–1512

44. Nopper AJ, Hoii KA, Sookdeo-Drost S, et al. Topical ointment therapy benefits premature infants. *J Pediatr.* 1996;128:660–669

45. Edwards WH, Conner JM, Soll RF, et al. The effect of prophylactic ointment therapy on nosocomial sepsis rates and skin integrity in infants with birth weights of 501 to 1000 g. *Pediatrics.* 2004;113:1195–1203

46. Soll RF, Morley CJ. Prophylactic versus selective use of surfactant in preventing morbidity and mortality in preterm infants. *Cochrane Database Syst Rev.* 2001;(2):CD000510

47. Subramaniam P, Henderson-Smart DJ, Davis PG. Prophylactic nasal continuous positive airways pressure for preventing morbidity and mortality in very preterm infants. *Cochrane Database Syst Rev.* 2005;(3):CD001243

48. Yost CC, Soll RF. Early versus delayed selective surfactant treatment for neonatal respiratory distress syndrome. *Cochrane Database Syst Rev.* 2000;(2):CD001456

49. Woodgate PG, Davies MW. Permissive hypercapnia for the prevention of morbidity and mortality in mechanically ventilated newborn infants. *Cochrane Database Syst Rev.* 2001;(2):CD002061

50. Askie LM, Henderson-Smart DJ. Restricted versus liberal oxygen exposure for preventing morbidity and mortality in preterm or low birth weight infants. *Cochrane Database Syst Rev.* 2001;(4):CD001077

51. Woodgate PG, Flenady V. Tracheal suctioning without disconnection in intubated ventilated neonates. *Cochrane Database Syst Rev.* 2001;(2):CD003065

52. Ng E, Taddio A, Ohlsson A. Intravenous midazolam infusion for sedation of infants in the neonatal intensive care unit. *Cochrane Database Syst Rev.* 2003;(1):CD002052

53. Shortell SM, Zimmerman JE, Rousseau DM, et al The performance of intensive care units: does good management make a difference? *Med Care.* 1994;32:508–525

54. Bell EF, Acarregui MJ. Restricted versus liberal water intake for preventing morbidity and mortality in preterm infants. *Cochrane Database Syst Rev.* 2001;(3):CD000503

55. McGuire W, Anthony MY. Formula milk versus preterm human milk for feeding preterm or low birth weight infants. *Cochrane Database Syst Rev.* 2001;(3):CD002972

A Cluster-Randomized Trial of Benchmarking and Multimodal Quality Improvement to Improve Rates of Survival Free of Bronchopulmonary Dysplasia for Infants With Birth Weights of Less Than 1250 Grams

Michele Walsh, Abbott Laptook, S. Nadya Kazzi, William A. Engle, Qing Yao, Maynard Rasmussen, Susie Buchter, Gregory Heldt, William Rhine, Rose Higgins and Kenneth Poole

Updated Information & Services	including high resolution figures, can be found at: http://pediatrics.aappublications.org/content/119/5/876
References	This article cites 42 articles, 9 of which you can access for free at: http://pediatrics.aappublications.org/content/119/5/876#BIBL
Subspecialty Collections	This article, along with others on similar topics, appears in the following collection(s): **Fetus/Newborn Infant** http://www.aappublications.org/cgi/collection/fetus:newborn_infant_sub
Permissions & Licensing	Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at: http://www.aappublications.org/site/misc/Permissions.xhtml
Reprints	Information about ordering reprints can be found online: http://www.aappublications.org/site/misc/reprints.xhtml

American Academy of Pediatrics
DEDICATED TO THE HEALTH OF ALL CHILDREN™

PEDIATRICS®

OFFICIAL JOURNAL OF THE AMERICAN ACADEMY OF PEDIATRICS

A Cluster-Randomized Trial of Benchmarking and Multimodal Quality Improvement to Improve Rates of Survival Free of Bronchopulmonary Dysplasia for Infants With Birth Weights of Less Than 1250 Grams

Michele Walsh, Abbott Laptook, S. Nadya Kazzi, William A. Engle, Qing Yao, Maynard Rasmussen, Susie Buchter, Gregory Heldt, William Rhine, Rose Higgins and Kenneth Poole

The online version of this article, along with updated information and services, is located on the World Wide Web at:

http://pediatrics.aappublications.org/content/119/5/876

American Academy of Pediatrics

DEDICATED TO THE HEALTH OF ALL CHILDREN™