Comparative Effectiveness and Practice Variation in Neonatal Care

Joanne Lagatta, MD, MS[a,*], Michael Uhing, MD[a],
Julie Panepinto, MD, MSPH[a,b]

KEYWORDS

- Comparative effectiveness • Neonatology • Clinical research methods

KEY POINTS

- Components of comparative effectiveness research (CER) include comparisons of alternative standards of care, evaluating outcomes important to individuals, and incorporating varied settings and participants.
- Neonatal clinical research contains examples of CER with strengths in clinical trials and metaanalyses comparing alternative standards of care.
- Future work in neonatal CER could focus on patient-centered outcomes in both prospective and retrospective studies.

INTRODUCTION

There is increasing discussion in medical literature and among grant funding agencies about the need for comparative effectiveness research (CER). CER is defined by the Institute of Medicine as "the generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor a clinical condition or to improve the delivery of care."[1] At first glance, this definition is broad enough that it potentially encompasses all types of clinical research, because the prevention, diagnosis, and treatment of illness is the ultimate goal of any clinical research team. In fact, neonatal clinical research literature already contains many examples of research that fit into the broad framework of CER. This article describes the main types of CER research methods using recent examples from existing neonatology literature, and highlights challenges in conducting CER specific to neonatal research.

WHAT IS COMPARATIVE EFFECTIVENESS RESEARCH?

The focus of CER is to assist patients, clinicians, and policymakers in making informed decisions to improve health care. Although a variety of research methods can be used to accomplish these goals, 4 key elements of CER have been identified.

Direct Comparison of Potential Alternative Standards of Care

In contrast with an efficacy trial of a novel intervention versus a placebo, an effectiveness study compares the outcome of at least 2 existing interventions that a patient or clinician could reasonably choose in day-to-day clinical practice. Specifically, an effectiveness study aims to determine if an intervention does work, whereas a traditional efficacy study aims to determine if it can work.[2]

Evaluating a Broad Array of Health-related Outcomes that Are Important to Individuals

Although there is often overlap, CER targets individual decision making by the patient or clinician, whereas public health decision making focuses on a population. CER studies assess both benefits and harms of interventions, and the measurement of clinical outcomes (as opposed to surrogate markers) that are important to decision-makers such as survival, daily functioning, symptoms, and health-related quality of life.

Incorporating a Wide Variety of Settings and Participants

In contrast with carefully controlled clinical trials evaluating a select group of patients, CER is meant to focus on a typical patient in a typical practice setting. In addition to estimating an average treatment effect across an entire study population, a goal of CER is to study heterogeneous effects within clinically relevant subgroups to help predict which individuals most benefit from treatment.[3]

Prioritizing Topics of Interest to Stakeholders

Involvement from patients, clinicians, policymakers, and other relevant participants in health care delivery is seen as key to guiding investment in research that reflects the priorities of the public.

HOW DOES COMPARATIVE EFFECTIVENESS RESEARCH RELATE TO QUALITY IMPROVEMENT?

Neonatal research has documented numerous examples of variation in neonatal care practices, such as use of inotropic agents,[4] use of home oxygen and diuretics for infants with bronchopulmonary dysplasia,[5,6] and antenatal counseling for preterm infants.[7] Likewise, there are numerous examples of variation in important clinical outcomes, such as mortality for extremely preterm infants,[8] bronchopulmonary dysplasia,[9] and length of hospital stay.[10] Quality improvement efforts focus on reducing variation in care and implementing best process and practice within individual neonatal intensive care units (NICUs) via education, monitoring changes in care and outcome, and benchmarking of outcomes against national standards.[11] For example, there have been multiple quality improvement efforts to reduce rates of bloodstream infections in the NICU in which multicenter groups work collaboratively to standardize and implement best practices identified through review of care at the centers with the lowest reported rates of bloodstream infections.[12–14]

CER, on the other hand, can be thought of as research directed at identifying best practices among the wide variation present in medical care. This is especially true when there are multiple treatments or processes available to treat the same condition.

In contrast to the quality improvement examples cited, a CER study on preventing bloodstream infection would compare outcomes after use of differing handwashing agents, types of catheters, or approaches to line placement, while using similar risk adjustment measures that are used for benchmarking outcomes in quality reporting.[11,15,16] For an outcome that is clearly positive or negative, such as line infection rates, the goal of CER is to provide evidence that can be implemented to optimize outcomes, and not necessarily directed at standardizing practice.

For outcomes in which family or clinician preferences vary, or optimal treatments differ by patient subgroup, a goal of CER is to provide enough evidence to allow clinicians and families to tailor their approach to achieve the outcomes that are best for the patient. In neonatology, tailored approaches to care based on parent preferences are discussed most commonly in the context of resuscitation for extremely preterm infants,[17] but could be applied to other medical decision making as well. For example, a NICU that allows more home nasogastric feedings, places gastrostomy tubes earlier, or prescribes home oxygen or apnea monitors more readily might have a shorter length of stay, but this shorter length of stay is not necessarily preferable to a family with more limited home resources. This is the rationale behind measuring a broad array of outcomes for any given intervention, as well as focusing on outcomes that are directly interpretable to families.

STUDY DESIGNS FOR COMPARATIVE EFFECTIVENESS RESEARCH

Any traditional clinical study design can be used to answer research questions that can be categorized as CER. Neonatal literature contains examples that highlight CER principles across categories of study designs and data sources. This article highlights 3 major categories: Randomized controlled trials (RCT), systematic reviews of existing literature, and observational studies. These examples are not meant as an exhaustive review of the literature, but rather to highlight existing neonatal research as examples of CER.

Randomized, Controlled Trials

RCTs are traditionally designed to evaluate the efficacy of therapies in a specific population of subjects, analyzing an intervention by comparing randomized groups to receive either a treatment or a blinded placebo under standardized conditions. RCTs are also used to determine effectiveness by comparing treatments or procedures. Successful randomization and blinding minimizes confounding, and is the reason that RCTs are considered the gold standard of clinical research.

The media controversy after the SUPPORT trial makes it seem as though effectiveness trials in neonatology are new,[18] but numerous neonatal RCTs are examples of effectiveness trials. Although comparisons of oxygen saturation targets are the most publicized recent clinical trials in this category,[19–21] many other studies fit this description. Trials comparing modes of ventilation for preterm infants can be considered comparative effectiveness trials, because there could be no ethical or practical placebo group receiving no respiratory support. Examples include prophylactic intubation and surfactant versus nasal continuous positive airway pressure (NCPAP) at resuscitation, or noninvasive ventilation comparisons, such as synchronized nasal intermittent positive pressure ventilation versus NCPAP or high-flow nasal cannula versus NCPAP.[22–24] Even in evaluations of drug therapies, neonatal literature has relevant examples, such as comparisons of bevacizumab versus laser therapy for retinopathy of prematurity,[25] dopamine versus epinephrine for hypotension,[26] and high versus low amino acid levels in parenteral nutrition for extremely low birth weight

infants.[27] Developing new therapies and testing them against a placebo is obviously crucial to major progress in care for critically ill infants; initial evaluations of surfactant and antenatal steroids led to enormous improvements in neonatal mortality and morbidity.[28,29] However, effectiveness trials such as optimizing the timing and preparation of surfactant and antenatal steroids,[30–32] and improving key components of neonatal intensive care management such as ventilation, blood pressure, and nutritional support, are needed to continue the significant improvements in mortality and morbidity seen since 2000.[33,34]

RCTs have limitations in the context of CER, which focuses on identifying the best treatment of available options individualized to relevant subgroups. "Standardized conditions" may have differing effects on the generalizability of trial results to a broader population. A larger sample size is often required to power a study comparing 2 treatment outcomes than to compare a single treatment to placebo, which makes enrollment more difficult. Inclusion criteria and the consent process may result in a different set of baseline characteristics between the trial cohort and the general population. This was noted most recently during enrollment for the SUPPORT trial, which noted differences between enrolled and nonenrolled patients by receipt of prenatal antibiotics, antenatal steroids, delivery room interventions, and outcomes including mortality, bronchopulmonary dysplasia (BPD) and severe intraventricular hemorrhage.[35,36] As a result, interpreting the study findings of higher mortality in the lower oxygen saturation target group has been difficult, because both groups had lower mortality than nonenrolled patients.[21,37,38] Measuring the effect of an intervention versus a placebo may require changes to "standard care" by altering the timing of the primary intervention and its associated care. Placebo effects may change the outcome in either a treatment or a control group. The process of selecting measurable outcome criteria with a limited sample size may or may not result in trial results that reflect the key factors used by a clinician in weighing treatment options.[39]

One criticism of RCTs in CER is that enrolled patients may be a narrow group without comorbidities. In neonatology, clinical trials often enroll patients in a stratified fashion based on gestational age (for diseases of prematurity) rather than excluding patients with comorbid conditions. Gestational age or birth weight groups are often the clinically relevant subgroup analysis. This was demonstrated in initial comparisons of prophylactic surfactant versus early NCPAP, which suggested that infants born at earlier gestational ages treated with early NCPAP may have higher incidence of pneumothorax,[40] which prompted more recent studies.[24]

Another potential limitation of RCTs for the purposes of CER is the site of care. Most patients in RCTs are treated in academic medical centers, and may have different baseline characteristics than patients treated in other settings. This may be less of an issue in neonatology than in other areas of medicine, such as primary care, because patients with high-risk conditions such as extreme prematurity or conditions requiring mechanical ventilation or surgical intervention are concentrated in level III/IV units. Because the enrollment criteria for many RCTs are based on gestational age, the study populations in academic and nonacademic centers are often comparable.[33,41] For patients with uncommon conditions requiring multiple subspecialty care, the only (and therefore typical practice environment) is an academic center. Although most RCTs are conducted at academic-affiliated centers, the Vermont Oxford Network has conducted trials that enroll patients within private centers, such as comparisons of ventilatory and heat loss prevention strategies during resuscitation of preterm infants.[30,42] For patients who do not require level III care, it would be of interest to compare outcomes for infants treated in lower acuity centers to be able to generalize results to the setting under which most of those infants are treated. As

with other fields of medicine, most lower acuity centers do not have the infrastructure or the volume to support prospective clinical trials.

Systematic Review and Metaanalysis

Systematic reviews and metaanalyses are overlapping methods of evaluating existing evidence. Systematic reviews use prespecified search methods to evaluate and synthesize eligible studies on a specific clinical question. Metaanalysis refers to a quantitative re-analysis of pooled data from individual studies.[43] These techniques allow results from multiple independent studies to be combined into a quantitative estimate of effect, such as combining results of multiple RCTs or epidemiologic studies. Neonatology has a strong history of systematic review, beginning with the Oxford Database of Perinatal Trials in the 1980s and continuing as the Cochrane Neonatal Group.[44,45] Numerous systematic reviews and metaanalyses are updated through the Cochrane Neonatal Reviews, which serve different goals within the framework of CER.[46,47] Many analyze direct comparisons of alternative potentially standard therapies, such as dopamine versus dobutamine for hypotensive preterm infants.[48] In addition, the synthesis of multiple similar studies can increase power to detect a treatment effect when not all individual studies have found statistical significance. They are advantageous in studying rare or adverse events, and can also highlight effects in relevant subgroups, which may be too small in single RCTs. The Cochrane review of ibuprofen for the treatment of patent ductus arteriosus provides an example of several of these advantages. Treatment with ibuprofen versus indomethacin showed equivalent effectiveness in patent ductus arteriosus closure, but less risk of necrotizing enterocolitis in pooled estimates, although no single trial showed a significant difference.[49] Although not strictly CER because it reviewed comparisons of an intervention with placebo, a significant example of neonatal research metaanalysis showing subgroup benefits was in the use of antenatal corticosteroids for prevention of respiratory distress syndrome. Crowley identified RDS reduction in infants less than 31 weeks by metaanalysis, although outcomes from individual studies were not significant.[50] Systematic reviews and metaanalyses can also report observational studies, although this is less common owing to differences in study design that limit pooled estimates.[51]

Observational Study Design in Comparative Effectiveness Research

The recent funding availability for CER has resulted in increased attention to observational studies. This raises controversy because of bias limitations inherent in observational study design. One of the most common problems with nonrandomized studies is the uneven distribution of unmeasured confounders. Another major issue is confounding by indication, meaning that the patients believed most likely to benefit from a treatment are the ones most likely to receive it, which exaggerates the actual treatment effect in the analysis. Time frames of the study cohort can present difficulties, with new entrants and attrition. Finally, practice and policy changes that occur during data collection can affect analysis. The data source can affect applicability to other settings, such as the characteristics of that population, local practice patterns, and resource availability.[52] Although quantitative methods to minimize the effect of bias are beyond the scope of this paper, the Institute of Medicine has recommended explicit attention to methodologic considerations of observational study design.[1,53]

Despite these limitations, observational studies provide a mechanism for answering clinical questions for which RCTs are not feasible for a variety of reasons. Observational studies offer potential benefit for clinical questions in which the required sample size would be prohibitive.[54] This could include evaluations of adverse events, such as comparing the effect of differing lengths of initial antibiotic therapy on subsequent

development of necrotizing enterocolitis.[55] Studies have found that confounding by indication is less problematic for evaluating unanticipated harms than for evaluating beneficial effects.[47] The large number of patients available for observational studies also facilitates studies of relatively rare diseases, such as the effect of antifungal therapy in extremely low birth weight infants with invasive candidiasis.[56] It can be useful in comparing the effects of similar therapies for which the sample size to detect treatment differences may be prohibitive, such as comparing types of antenatal corticosteroids on subsequent hearing and neurodevelopmental impairment.[57] Observational studies are also important when randomization is not feasible owing to ethical considerations, or practical issues related to the study question at hand, such as questions of treatment adherence or usage outside of trials, such as during evaluation of total body cooling for hypoxic ischemic encephalopathy,[58] or geographic or demographic effects on treatment results.[54] Finally, observational studies, particularly those using already existing data, are far less expensive and time consuming than RCTs. For clinical questions where an RCT would be cost or time prohibitive, observational data represent an alternative to expedite advancing the evidence basis for clinical decision making.[54]

Data Sources for Observational Study Design

Observational studies can be conducted as a prospective cohort, as was done in comparing antihypotensive therapies for extremely preterm infants.[59] In this example, an observational design was chosen over an RCT owing to lack of physician equipoise in treating hypotension, wide practice variability complicating identification of inclusion and exclusion criteria, and the potential for enrollment or selection bias when enrolling a vulnerable patient population shortly after birth.[35,59] More commonly, to obtain the sample size that confers an advantage to observational studies, they can be accomplished via secondary analysis of already collected data. Data sources include disease registries, electronic health records data, and administrative data.

Very low birth weight registries such as the National Institute of Child Health and Human Development Neonatal Research Network and Vermont Oxford Network, or disease registries such as the Extracorporeal Life Support Organization database and the Congenital Diaphragmatic Hernia Study Group, are gathered via primary data abstraction of prespecified data elements. Benefits to this approach include trained abstractors, ongoing quality assurance regarding data collection, and discrete coded variables that relate to the disease of interest. Limitations to this approach include lack of granularity of the data fields, differences in definitions between data sources, and differences in the way that detailed individual data (eg, serial laboratory values) are aggregated into discrete variables. For example, the Vermont Oxford Network collects information on an infant's respiratory support at 36 weeks postmenstrual age. This type of data does then not allow comparison of therapies that require the need to discriminate between the differences in respiratory requirements that may be relevant only before or after this particular data point.

Following the US government meaningful use incentives for use of electronic health records systems, chart review for clinical research is rapidly becoming a more viable large-scale option.[60] Secondary analyses of clinical data can be more easily facilitated in single centers as well as in emerging collaborative arrangements.[61] Detailed clinical data are available, although accurate abstraction depends on consistency in documenting the start and resolution of illnesses, and limiting diagnostic variability between centers for common illnesses such as apnea.[62,63] In addition to difficulties inherent in obtaining research data from a clinical chart, challenges to study design include

identification of interventions and outcomes, a study population and follow-up interval, and a plan for active versus passive data capture.[64] One of the most established sources of secondary analysis using medical records data in neonatology comes from the Pediatrix Medical Group, whose Clinical Data Warehouse automatically facilitates export of de-identified, discrete data elements from patient charts.[65] Several comparative effectiveness studies have come from this group, including comparisons of adverse events after differing preparations of surfactant[32] and different empiric antibiotic regimens.[66]

In many fields of medicine, administrative data taken from billing claims are commonly used for CER. Compared with patient registry or electronic health record data, these data sources are more likely to be limited by lack of completeness of the listed diagnoses. Acute and particularly surgical conditions are more likely to be coded appropriately. Neonatology, as in many other fields, has some discrepancies between clinical diagnoses and coding terminology for common diseases, such as bronchopulmonary dysplasia and respiratory distress syndrome, which limits the clinical detail needed to design CER studies.[67] Public use data files, such the National Inpatient Sample and its associated Kids Inpatient Database available from the Agency for Healthcare Research and Quality, have similar limitations in diagnosis availability because they are derived from billing data. However, billing files are a potential source of post-NICU follow-up data on rehospitalizations or costs of outpatient care, which could provide useful outcome measures. Linked data sets that pair the longitudinal data collection of billing data with the appropriate amount of neonatal coding accuracy, such as the California Perinatal Quality Care Collaborative,[68] the Kaiser Permanente Neonatal Minimum Data Set,[69] or the Children's Hospitals Association Neonatal Database,[63] could be used to conduct this type of research.

CHALLENGES IN NEONATAL COMPARATIVE EFFECTIVENESS RESEARCH
Comparing Treatment with Placebo: Efficacy Versus Effectiveness

Perhaps more commonly than in other clinical fields of medicine, for many diseases of prematurity, a potential "standard of care" is observation without intervention. There are no noted differences in mortality or morbidity whether a hypotensive extremely low birth weight infant with reasonable end-organ perfusion receives a vasopressor or clinical observation alone.[59] Thus, in some circumstances, comparing an intervention with observation alone could be considered an effectiveness study. Historically, some treatments that were adopted as standard practice without controlled trials, such as bicarbonate for metabolic acidosis, turned out to be worse than placebo.[70] Although there are other instances in medicine where observation alone is a viable treatment option (early stage prostate cancer), very few other specialties see an individual patient grow 5- to 7-fold over the course of a single hospitalization, making observation without intervention potentially a more relevant therapeutic option in neonatology than in other fields.

Defining and Measuring Patient-Centered Outcomes

The Institute of Medicine has identified "patient-centered outcomes" as outcomes that are directly relevant to stakeholders, rather than proxy measures. Depending on the research question, a stakeholder could be a patient or parent, a practicing clinician, or a health system administrator developing practice standards. In outlining standards for patient-centered research, the Patient Centered Outcomes Research Institute encourages providing information supporting the selection of outcomes as clinically meaningful, such as input from patients and their families.

Deciding which stakeholder's perspective drives outcome selection makes a difference in study design. For example, leaders of a NICU wishing for their hospital to compare favorably in outcomes reporting across the Vermont Oxford Network or their multi-unit practice group may be interested in strategies to reduce their unit's rate of BPD. However, families may be less concerned about whether their infant requires oxygen at 36 weeks' postmenstrual age, versus whether their infant requires home oxygen at discharge or requires rehospitalization after leaving the NICU. When a reduction in a proximal morbidity such as BPD results in a better long-term health outcome, all stakeholders are mutually satisfied. However, a recent publication from the SUP-PORT trial highlights the fact that proximal and distal outcomes are not necessarily equivalent: The study found no difference in the primary composite outcome of death or bronchopulmonary dysplasia, but significant differences in more functional outcomes, such as wheezing.[71] In this case, it seems straightforward for a clinician to choose the treatment option that results in better patient outcomes. But functional outcomes such as wheezing, or even health care utilization measures such as readmissions after NICU discharge, are more variably defined and reported than proximal morbidities such as BPD.[72] Obtaining consensus on best treatment strategies based on patient-centered outcomes requires more effort in selecting and defining outcomes that matter to a broader range of stakeholders.

In general, designing and powering studies that measure patient-centered outcomes in neonatal CER will continue to be a challenge. RCTs are often powered for a set of primary endpoints, such as death or BPD, and measure multiple secondary endpoints, such as respiratory symptoms. Interpreting results of multiple secondary outcomes is difficult when adjusting for multiple comparisons. Involving families explicitly in study design could provide additional insights into selection of outcomes from the multiple available options. In the United Kingdom, many health research funding bodies require patient involvement in study design. Consumer involvement has been reported in convincing funders of study relevance, developing clear and relevant study questions and outcomes, providing insights into patients' views of trial logistics, advising on recruitment, and developing patient information materials.[73,74] Neonatal research has noted differences in parent and clinician perspectives about neonatal treatments for extremely preterm infants and infants with trisomy 13 and 18[75,76]; differing perspectives are also likely present and insightful for decision making that involves less life-and-death situations.

If the large sample size available for observational studies could facilitate evaluation of multiple outcomes, it could be easier to evaluate enough patient-centered outcomes to appeal to multiple stakeholders. Currently, many observational data sources lack discrete data on patient-centered outcomes besides mortality and length of stay. Measuring symptoms such as pain and fatigue rather than laboratory values or definitions (BPD) presents unique issues in a very young pediatric population because they require either clinician assessment or parent report, and there are no validated multidomain quality of life measures for infants under 1 month of age. Two-year neurodevelopmental assessments are often used as proxy measures for long-term functioning, although the correlation between 2-year assessments and later functioning is imperfect.[77] There are many longitudinal studies of quality of life in survivors of preterm birth,[78] but none have been used yet for CER. However, this could be an area for future development in neonatal patient-centered outcomes research. Nurses already record discrete pain scores, and therapists perform standardized motor assessments, which are part of the health record and could be abstracted for observational research.[79] For infants older than 1 month of age there are validated tools for infants and young children, as well as for caregivers related to their child's illness.[80,81] In adult

populations with wider use of patient-reported outcome measures, such as the National Institutes of Health-supported Patient-reported Outcomes Measurement Information System (PROMIS) measures, there are increasing efforts to embed patient-reported outcomes assessments into the electronic health record for clinical and research purposes.[82,83] There is a need for more work on defining key patient-centered measures for infants in the NICU and in postdischarge follow-up to help focus efforts in neonatal CER.

SUMMARY

CER is a relatively new focus area that encompasses principles of clinical research that to some extent have been present in neonatal research for some time. Identifying best practices within the wide variation of NICU care can enable clinicians and quality improvement efforts to either standardize care across groups, or tailor efforts to specific patients that benefit from particular approaches. This increased emphasis on evidence to direct decision making ties in well with neonatology's already strong efforts in quality improvement, systematic reviews, patient registries and multicenter efforts. Many data sources exist with the potential to increase CER efforts within neonatology. Future work is needed to define patient-centered outcomes to focus prospective clinical studies, and embed appropriate tools within observational data to facilitate analysis of patient-centered outcomes.

REFERENCES

1. Sox HC, Greenfield S. Comparative effectiveness research: a report from the Institute of Medicine. Ann Intern Med 2009;151:203–5.
2. Luce BR, Drummond M, Jonsson B, et al. EBM, HTA, and CER: clearing the confusion. Milbank Q 2010;88:256–76.
3. Varadhan R, Segal JB, Boyd CM, et al. A framework for the analysis of heterogeneity of treatment effect in patient-centered outcomes research. J Clin Epidemiol 2013;66:818–25.
4. Wong J, Shah PS, Yoon EW, et al. Inotrope use among extremely preterm infants in Canadian neonatal intensive care units: variation and outcomes. Am J Perinatol 2014. [Epub ahead of print].
5. Lagatta J, Clark R, Spitzer A. Clinical predictors and institutional variation in home oxygen use in preterm infants. J Pediatr 2012;160:232–8.
6. Slaughter JL, Stenger MR, Reagan PB. Variation in the use of diuretic therapy for infants with bronchopulmonary dysplasia. Pediatrics 2013;131:716–23.
7. Mehrotra A, Lagatta J, Simpson P, et al. Variations among US hospitals in counseling practices regarding prematurely born infants. J Perinatol 2013;33:509–13.
8. Alleman BW, Bell EF, Li L, et al. Individual and center-level factors affecting mortality among extremely low birth weight infants. Pediatrics 2013;132:e175–84.
9. Ambalavanan N, Walsh M, Bobashev G, et al. Intercenter differences in bronchopulmonary dysplasia or death among very low birth weight infants. Pediatrics 2011;127:e106–16.
10. Murthy K, Dykes FD, Padula MA, et al. The Children's Hospitals neonatal database: an overview of patient complexity, outcomes and variation in care. J Perinatol 2014;34(8):582–6.
11. Horbar JD. The Vermont Oxford Network: evidence-based quality improvement for neonatology. Pediatrics 1999;103:350–9.

12. Schulman J, Stricof R, Stevens TP, et al. Statewide NICU central-line-associated bloodstream infection rates decline after bundles and checklists. Pediatrics 2011;127:436–44.
13. Aly H, Herson V, Duncan A, et al. Is bloodstream infection preventable among premature infants? A tale of two cities. Pediatrics 2005;115:1513–8.
14. Kilbride HW, Wirtschafter DD, Powers RJ, et al. Implementation of evidence-based potentially better practices to decrease nosocomial infections. Pediatrics 2003;111:e519–33.
15. Rogowski JA, Horbar JD, Staiger DO, et al. Indirect vs direct hospital quality indicators for very low-birth-weight infants. JAMA 2004;291:202–9.
16. Richardson D, Tarnow-Mordi WO, Lee SK. Risk adjustment for quality improvement. Pediatrics 1999;103:255–65.
17. Mercurio MR. Parental authority, patient's best interest and refusal of resuscitation at borderline gestational age. J Perinatol 2006;26:452–7.
18. Lantos JD. Learning the right lessons from the SUPPORT study controversy. Arch Dis Child Fetal Neonatal Ed 2014;99:F4–5.
19. Group BIUKC, Group BIAC, Group BINZC, et al. Oxygen saturation and outcomes in preterm infants. N Engl J Med 2013;368:2094–104.
20. Schmidt B, Whyte RK, Asztalos EV, et al. Effects of targeting higher vs lower arterial oxygen saturations on death or disability in extremely preterm infants: a randomized clinical trial. JAMA 2013;309:2111–20.
21. SUPPORT Study Group of the Eunice Kennedy Shriver NICHD Neonatal Research Network, Carlo WA, Finer NN, Walsh MC, et al. Target ranges of oxygen saturation in extremely preterm infants. N Engl J Med 2010;362:1959–69.
22. Bhandari V, Finer NN, Ehrenkranz RA, et al. Synchronized nasal intermittent positive-pressure ventilation and neonatal outcomes. Pediatrics 2009;124:517–26.
23. Yoder BA, Stoddard RA, Li M, et al. Heated, humidified high-flow nasal cannula versus nasal CPAP for respiratory support in neonates. Pediatrics 2013;131:e1482–90.
24. Early CPAP versus surfactant in extremely preterm infants. N Engl J Med 2010;362:1970–9.
25. Mintz-Hittner HA, Kennedy KA, Chuang AZ. Efficacy of intravitreal bevacizumab for stage 3+ retinopathy of prematurity. N Engl J Med 2011;364:603–15.
26. Valverde E, Pellicer A, Madero R, et al. Dopamine versus epinephrine for cardiovascular support in low birth weight infants: analysis of systemic effects and neonatal clinical outcomes. Pediatrics 2006;117:e1213–22.
27. Burattini I, Bellagamba MP, Spagnoli C, et al. Targeting 2.5 versus 4 g/kg/day of amino acids for extremely low birth weight infants: a randomized clinical trial. J Pediatr 2013;163:1278–82.e1.
28. Seger N, Soll R. Animal derived surfactant extract for treatment of respiratory distress syndrome. Cochrane Database Syst Rev 2009;(2):CD007836.
29. Roberts D, Dalziel S. Antenatal corticosteroids for accelerating fetal lung maturation for women at risk of preterm birth. Cochrane Database Syst Rev 2006;(3):CD004454.
30. Dunn MS, Kaempf J, de Klerk A, et al. Randomized trial comparing 3 approaches to the initial respiratory management of preterm neonates. Pediatrics 2011;128:e1069–76.
31. Guinn DA, Atkinson MW, Sullivan L, et al. Single vs weekly courses of antenatal corticosteroids for women at risk of preterm delivery: a randomized controlled trial. JAMA 2001;286:1581–7.

32. Trembath A, Hornik CP, Clark R, et al. Comparative effectiveness of surfactant preparations in premature infants. J Pediatr 2013;163:955–60.e1.
33. Horbar JD, Carpenter JH, Badger GJ, et al. Mortality and neonatal morbidity among infants 501 to 1500 grams from 2000 to 2009. Pediatrics 2012;129: 1019–26.
34. Fanaroff AA, Stoll BJ, Wright LL, et al. Trends in neonatal morbidity and mortality for very low birthweight infants. Am J Obstet Gynecol 2007;196:147.e1-8.
35. Rich W, Finer NN, Gantz MG, et al. Enrollment of extremely low birth weight infants in a clinical research study may not be representative. Pediatrics 2012;129:480–4.
36. Rich WD, Auten KJ, Gantz MG, et al. Antenatal consent in the SUPPORT trial: challenges, costs, and representative enrollment. Pediatrics 2010;126:e215–21.
37. Bancalari E, Claure N. Oxygenation targets and outcomes in premature infants. JAMA 2013;309:2161–2.
38. Polin RA, Bateman D. Oxygen-saturation targets in preterm infants. N Engl J Med 2013;368:2141–2.
39. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. J Clin Epidemiol 2009;62:499–505.
40. Morley CJ, Davis PG, Doyle LW, et al. Nasal CPAP or intubation at birth for very preterm infants. N Engl J Med 2008;358:700–8.
41. Stoll BJ, Hansen NI, Bell EF, et al. Neonatal outcomes of extremely preterm infants from the NICHD neonatal research network. Pediatrics 2010;126:443–56.
42. Vohra S, Reilly M, Rac VE, et al. Study protocol for multicentre randomized controlled trial of HeLP (Heat loss prevention) in the delivery room. Contemp Clin Trials 2013;36:54–60.
43. Akobeng AK. Understanding systematic reviews and meta-analysis. Arch Dis Child 2005;90:845–8.
44. Chalmers I, Hetherington J, Newdick M, et al. The Oxford database of perinatal trials: developing a register of published reports of controlled trials. Control Clin Trials 1986;7:306–24.
45. Cochrane Neonatal Group. Available at: http://neonatal.cochrane.org/.
46. Normand SL. Meta-analysis: formulating, evaluating, combining, and reporting. Stat Med 1999;18:321–59.
47. Chou R, Helfand M. Challenges in systematic reviews that assess treatment harms. Ann Intern Med 2005;142:1090–9.
48. Subhedar NV, Shaw NJ. Dopamine versus dobutamine for hypotensive preterm infants. Cochrane Database Syst Rev 2003;(3):CD001242.
49. Ohlsson A, Walia R, Shah SS. Ibuprofen for the treatment of patent ductus arteriosus in preterm and/or low birth weight infants. Cochrane Database Syst Rev 2013;(4):CD003481.
50. Crowley P. Prophylactic corticosteroids for preterm birth. Cochrane Database Syst Rev 2006;(2):CD000065.
51. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. meta-analysis of observational studies in epidemiology (MOOSE) group. JAMA 2000;283:2008–12.
52. Cox E, Martin BC, Van Staa T, et al. Good research practices for comparative effectiveness research: approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects using secondary data sources: the international society for pharmacoeconomics and outcomes research good research practices for retrospective database analysis task force report—part II. Value Health 2009;12:1053–61.

53. Helfand M, Tunis S, Whitlock EP, et al. A CTSA agenda to advance methods for comparative effectiveness research. Clin Transl Sci 2011;4:188–98.

54. Dreyer NA, Tunis SR, Berger M, et al. Why observational studies should be among the tools used in comparative effectiveness research. Health Aff (Millwood) 2010;29:1818–25.

55. Cotten CM, Taylor S, Stoll B, et al. Prolonged duration of initial empirical antibiotic treatment is associated with increased rates of necrotizing enterocolitis and death for extremely low birth weight infants. Pediatrics 2009;123:58–66.

56. Greenberg RG, Benjamin DK Jr, Gantz MG, et al. Empiric antifungal therapy and outcomes in extremely low birth weight infants with invasive candidiasis. J Pediatr 2012;161:264–9.e2.

57. Lee BH, Stoll BJ, McDonald SA, et al. Neurodevelopmental outcomes of extremely low birth weight infants exposed prenatally to dexamethasone versus betamethasone. Pediatrics 2008;121:289–96.

58. Azzopardi D, Strohm B, Edwards AD, et al. Treatment of asphyxiated newborns with moderate hypothermia in routine clinical practice: how cooling is managed in the UK outside a clinical trial. Arch Dis Child Fetal Neonatal Ed 2009;94: F260–4.

59. Batton B, Li L, Newman NS, et al. Use of antihypotensive therapies in extremely preterm infants. Pediatrics 2013;131:e1865–73.

60. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N Engl J Med 2010;363:501–4.

61. Narus SP, Srivastava R, Gouripeddi R, et al. Federating clinical data from six pediatric hospitals: process and initial results from the PHIS+ Consortium. AMIA Annu Symp Proc 2011;2011:994–1003.

62. Eichenwald EC, Zupancic JA, Mao WY, et al. Variation in diagnosis of apnea in moderately preterm infants predicts length of stay. Pediatrics 2011;127:e53–8.

63. Pallotto EK, Hunt PG, Dykes FD, et al. Topics in neonatal informatics: infants and data in the electronic health record era. Neoreviews 2013;14:e57–62.

64. Berger ML, Mamdani M, Atkins D, et al. Good research practices for comparative effectiveness research: defining, reporting and interpreting nonrandomized studies of treatment effects using secondary data sources: the ISPOR good research practices for retrospective database analysis task force report—part I. Value Health 2009;12:1044–52.

65. Spitzer AR, Ellsbury DL, Handler D, et al. The Pediatrix baby steps data warehouse and the Pediatrix quality steps improvement project system–tools for "meaningful use" in continuous quality improvement. Clin Perinatol 2010;37: 49–70.

66. Clark RH, Bloom BT, Spitzer AR, et al. Empiric use of ampicillin and cefotaxime, compared with ampicillin and gentamicin, for neonates at risk for sepsis is associated with an increased risk of neonatal death. Pediatrics 2006;117:67–74.

67. Pallotto EK, Hunt PG, Reber K, et al. Topics in neonatal informatics: standardizing diagnoses in neonatology: bronchopulmonary dysplasia and beyond. Neoreviews 2012;13:e577–82.

68. Gould JB. The role of regional collaboratives: the California perinatal quality care collaborative model. Clin Perinatol 2010;37:71–86.

69. Escobar GJ, Fischer A, Kremers R, et al. Rapid retrieval of neonatal outcomes data: the Kaiser Permanente neonatal minimum data set. Qual Manag Health Care 1997;5:19–33.

70. Aschner JL, Poland RL. Sodium bicarbonate: basically useless therapy. Pediatrics 2008;122:831–5.

71. Stevens TP, Finer NN, Carlo WA, et al. Respiratory outcomes of the surfactant positive pressure and oximetry randomized trial (SUPPORT). J Pediatr 2014; 165(2):240–9.e4.

72. Jobe AH, Bancalari E. Bronchopulmonary dysplasia. Am J Respir Crit Care Med 2001;163:1723–9.

73. Hanley B, Truesdale A, King A, et al. Involving consumers in designing, conducting, and interpreting randomised controlled trials: questionnaire survey. BMJ 2001;322:519–23.

74. Evans BA, Bedson E, Bell P, et al. Involving service users in trials: developing a standard operating procedure. Trials 2013;14:219.

75. Janvier A, Farlow B, Wilfond BS. The experience of families with children with trisomy 13 and 18 in social networks. Pediatrics 2012;130(2):293–8.

76. Partridge JC, Martinez AM, Nishida H, et al. International comparison of care for very low birth weight infants: parents' perceptions of counseling and decision-making. Pediatrics 2005;116:e263–71.

77. Hack M, Taylor HG, Drotar D, et al. Poor predictive validity of the Bayley scales of infant development for cognitive function of extremely low birth weight children at school age. Pediatrics 2005;116:333–41.

78. Saigal S, Stoskopf B, Pinelli J, et al. Self-perceived health-related quality of life of former extremely low birth weight infants at young adulthood. Pediatrics 2006; 118:1140–8.

79. Boss RD, Kinsman HI, Donohue PK. Health-related quality of life for infants in the neonatal intensive care unit. J Perinatol 2012;32:901–6.

80. Varni JW, Limbers CA, Neighbors K, et al. The PedsQL infant scales: feasibility, internal consistency reliability, and validity in healthy and ill infants. Qual Life Res 2011;20:45–55.

81. Varni JW, Sherman SA, Burwinkle TM, et al. The PedsQL family impact module: preliminary reliability and validity. Health Qual Life Outcomes 2004;2:55.

82. PROMIS: Dynamic Tools to Measure Health Outcomes from the Patient Perspective. Available at: http://www.nihpromis.org/.

83. Wu AW, Kharrazi H, Boulware LE, et al. Measure once, cut twice—adding patient-reported outcome measures to the electronic health record for comparative effectiveness research. J Clin Epidemiol 2013;66:S12–20.