Published in final edited form as:

Semin Perinatol. 2016 October ; 40(6): 374–384. doi:10.1053/j.semperi.2016.05.005.

Methodological Issues in the Design and Analyses of Neonatal Research Studies: Experience of the NICHD Neonatal Research Network

Abhik Das, PhD[1], **Jon Tyson, MD, MPH**[2], **Claudia Pedroza, PhD**[2], **Barbara Schmidt, MD, MSc**[3], **Marie Gantz, PhD**[4], **Dennis Wallace, PhD**[4], **William E. Truog, MD**[5], and **Rosemary D. Higgins, MD**[6]

[1]Biostatistics and Epidemiology Division, RTI International, Rockville, MD

[2]University of Texas Health Science Center at Houston, Houston, TX

[3]Department of Pediatrics, University of Pennsylvania, Philadelphia, PA

[4]Biostatistics and Epidemiology Division, RTI International, Research Triangle Park, NC

[5]Children's Mercy Hospitals and Clinics and the University of Missouri-Kansas City School of Medicine, Kansas City, MO

[6]Eunice Kennedy Shriver National Institute of Health & Human Development, National Institutes of Health, Bethesda, MD

Abstract

Objective—Impressive advances in neonatology have occurred over the 30 years of life of The Eunice Kennedy Shriver National Institute of Child Health and Human Development Neonatal Research Network (NRN). However, substantial room for improvement remains in investigating and further developing the evidence base for improving outcomes among the extremely premature. We discuss some of the specific methodological challenges in the statistical design and analysis of randomized trials and observational studies in this population.

Findings—Challenges faced by the NRN include designing trials for unusual or rare outcomes, accounting for and explaining center variations, identifying other subgroup differences, and balancing safety and efficacy concerns between short-term hospital outcomes and longer term neurodevelopmental outcomes.

Conclusions—The constellation of unique patient characteristics in neonates calls for broad understanding and careful consideration of the issues identified in this paper for conducting rigorous studies in this population.

Send all correspondence to: Abhik Das, PhD, RTI International, 6110 Executive Blvd., Suite 902, Rockville, MD 20852, Voice: (301) 770-8214, Fax: (301) 230-4646, adas@rti.org.

INTRODUCTION

Impressive advances in neonatology have occurred over the 30 years of life of The Eunice Kennedy Shriver National Institute of Child Health and Human Development Neonatal Research Network (NRN). However, improvement in survival for extremely premature babies has plateaued in recent years despite more aggressive use of antenatal steroids, antibiotics, and surfactant, while in-hospital morbidities, such as bronchopulmonary dysplasia (BPD), retinopathy of prematurity (ROP), intracranial hemorrhage (ICH), and sepsis remain high[1–3] as premature births have increased.[4] Although this has led to an increase in the number of infants at higher risk for long-term neurodevelopmental impairment (NDI), most investigators have reported that rates of neurologically intact survival (among live births) at 18 months to 2 years remain unchanged. Allowing for the complexities in interpreting a composite outcome (survival free of NDI) discussed later, overall this illustrates both the progress made in improving outcomes and the substantial room for improvement that remains.[5,6] These trends collectively have widespread implications for health care delivery and point to the continued need for targeted and rigorous research to develop better treatment and management strategies for neonates that improve long-term rehabilitation and outcome.

Reducing the high rates of in-hospital morbidity and later NDI among extremely premature infants remains a significant public health challenge, highlighting the importance of ongoing evidence-based research in this area. In addition, although late preterm births currently account for 75% of all neonatal intensive care unit (NICU) admissions, little evidence-based research occurs for these infants.[7,8] Thus, critical gaps in neonatal research remain, and many of the more severe diseases (such as necrotizing enterocolitis [NEC] or neonatal encephalopathy) are relatively infrequent conditions that require multicenter involvement to study them. Significant methodological challenges also exist in designing rigorous trials for unusual or rare outcomes, accounting for and explaining center variations, identifying other subgroup differences, and balancing safety and efficacy concerns between short-term hospital outcomes and longer term neurodevelopmental outcomes. These challenges require innovative trial design and analysis strategies to address them. Over the past 30 years of its existence, the NRN has conducted important studies to fill critical evidence gaps in the field and tackled several methodological issues in study design and data analyses in this area (see Table 1). The following discussion briefly highlights some of the special methodological concerns in neonatal studies and the NRN experience in addressing them.

STATISTICAL CONSIDERATIONS FOR NEONATAL RESEARCH

Designing studies in neonatal populations involves particular challenges, some of which apply almost universally across all NRN studies (e.g., the need to balance both proximal and distal outcomes to evaluate safety and efficacy), and some of which arise during the planning of specific studies (e.g., switching drug administration mode from IV to oral as an infant matures in a pharmacokinetics [PK] study). In this section, we give examples of statistical innovations we have used to overcome specific challenges in recent NRN studies, and we describe methods emerging from the research of statisticians at RTI and elsewhere that can be applied to challenges we foresee for future studies. The approaches discussed here

augment standard statistical techniques used in more straightforward studies and analyses, which are not discussed here.

Competing Outcomes

An important issue in designing randomized clinical trials (RCTs) for high-risk patients is the selection of an appropriate primary outcome when death is a competing outcome. In this situation, some patients will die before the outcome that the intervention is expected to prevent can be diagnosed, e.g. ROP. For this reason, the primary outcome in such trials is often a composite outcome, e.g., Death or BPD even if the intervention is not expected to affect mortality.

As illustrated by the Network SUPPORT trial, an important advantage of including deaths in the primary outcome is that mortality might unexpectedly be affected. This trial assessed whether use of an oxygen saturation goal in the lower half of the recommended range would reduce severe ROP among infants born at 24 to 27 weeks gestation. Based on the best available evidence before the trial, no effect on mortality was hypothesized or expected by the investigators, contrary to what critics unfamiliar with the issue of competing outcomes have assumed about this study.[9] The lower saturation goal did reduce severe ROP. However, this benefit was offset by an unexpected increase in death with no significant effect on the primary outcome of death or severe ROP. Had the primary outcome been severe ROP alone (among survivors), the primary outcome would not have captured the most important effect of the lower saturation goal, and the finding a mortality difference that prompted the recommendation to use a high saturation goal.[10]

In general, unless death is part of the primary outcome for a trial in a population that is likely to experience a substantially high death rate before the true outcome of interest can be assessed (again, because death is a competing outcome for any later morbidity in the NICU population), interpreting downstream events is complicated and inferences about the effects of treatment on these risks may be biased. This is because differential death rates in the comparison groups makes the survivors in these two groups a non-random sample of the randomized population, resulting in a biased and non-randomized comparison of outcome rates.

In addition to accounting for mortality, certain composite outcomes in neonatal research have considerable public health significance in their own right. Since the survival of extremely premature babies with profound impairment often has lifelong significance for these children, their families and society at large, survival free of neurodevelopmental impairment (or, neurologically intact survival) is a clinically meaningful outcome in its own right that is frequently used as a primary outcome for many NRN trials. For example, the ongoing Transfusion of Prematures (TOP) trial aims to examine whether the clinically relevant composite primary outcome of death or significant neurodevelopmental impairment in survivors at 22–26 months of corrected age is less common among preterm infants who, by transfusion practice, are maintained at higher hemoglobin levels.[11]

Although competing outcomes are an unavoidable problem, some investigators resist the use of composite outcomes, including those that include death as a competing outcome, because

outcomes of differing importance are given equal weight. In principle, this is not an inherent problem for composite outcomes. A potential solution is to weight the different outcomes according to their importance (utility) as judged by patients (or their family members) who know the most about or are most likely to experience these outcomes.[12–14] However difficult this approach may seem, in the long run it is likely to be necessary to resolve not only how different components of a composite outcome should be weighted but also the broader problem of determining when the benefits of treatment outweigh the hazards.

Accounting for competing and composite outcomes in primary outcome selection, sample size/power calculations, and interim monitoring plans is often necessary for neonatal trials. The NRN has helped develop innovative approaches to this issue, illustrated by the design of a current RCT to evaluate the use of hydrocortisone to reduce BPD.[15] The trial seeks to test whether a 10-day tapering course of hydrocortisone treatment for infants <30 weeks estimated gestational age at birth who remain intubated at 14–28 days postnatal age can reduce the composite outcome of BPD or death at 36 weeks PMA. However, because of safety concerns for hydrocortisone, particularly with respect to neurodevelopment, the intervention will not be considered successful if it increased the rates of NDI or death at follow-up. Thus, the primary outcome for this study includes both a measure of efficacy (improvement in survival without physiologically defined moderate to severe BPD) and safety (survival without moderate or severe NDI at 22–26 months corrected age). Consequently, we designed the study to provide a sequential evaluation of the composite hypothesis that (i) administration of hydrocortisone has efficacy benefit in that it reduces the risk of death or BPD at 36 weeks, and (ii) administration of hydrocortisone has an acceptable long-term safety profile with respect to death or NDI at 22–26 months.

This composite hypothesis will be evaluated sequentially. First, we will test the hypothesis that infants treated with hydrocortisone have a lower risk of death or BPD at 36 weeks than infants treated with placebo. If this test of efficacy indicates benefit of the hydrocortisone arm, then the safety of the treatment will be evaluated descriptively through an assessment of the risk to benefit ratio of the hydrocortisone treatment. Specifically, this safety outcome will be considered a “success” if either (1) the point estimate of risk of death/NDI is lower on the hydrocortisone arm than on the control arm, or (2) there is an increase in risk on the hydrocortisone arm, but the lower limit of a one-sided 95% confidence interval for the ratio of increased benefit for BPD to increased risk for NDI is greater than 4. In other words, for every additional four infants surviving without BPD, no more than one would experience death/NDI. Sample size calculations for this trial accounted for this design feature by first using standard power analysis to assess the sample size required to demonstrate a difference in death or BPD at 36 weeks on the two treatment arms. Second, simulation analyses were used to assess the probability of success for the safety outcome based on either of the two criteria defined above. The empirical probability of success for the descriptive analyses used to evaluate the safety outcome is analogous to the power of a formal hypothesis test.

Although the most common primary outcome in NRN studies is death or NDI, we anticipate future studies in which NDI or cognitive performance among survivors may be of primary interest. One possible analytical approach is to treat deaths as informatively missing and compare treatments within principal strata defined by potential survival outcomes.[16] As

appropriate, other emerging methods for the analysis of semicompeting risks such as semiparametric regression models,[17] illness–death models with shared frailty,[18] and multistate modeling in which individuals pass through various states (e.g., impairment and death, with interval censoring to accommodate preplanned evaluation times)[19] may be considered.

Trials for Conditions with Low Prevalence

Neonatal conditions such as NEC, severe respiratory failure requiring extracorporeal membrane oxygenation, and perinatal hypoxiaischemia have low prevalence but severe consequences. Although RCTs are the gold standard in clinical research, attaining sufficient enrollment for an adequately powered RCT in a reasonable amount of time can be challenging, and randomizing patients is not always feasible. A number of approaches have been used by NRN trials to address this recurring issue.

1. Comprehensive Cohort Design. One approach used in the ongoing NEC Surgery Trial comparing death or NDI for treatment with initial drain versus initial laparotomy in extremely premature babies with NEC was to implement a comprehensive cohort design[20] to include both randomized and nonrandomized (physician preference) patients, with causal inference techniques used for the latter.[21,22] Comprehensive cohort designs may be useful in judging whether treatment effects on the risk-adjusted outcomes of eligible nonrandomized patients are comparable to those among randomized patients. However, this assessment requires that all or virtually all eligible nonrandomized infants, including those who die, be included to avoid selection bias. Ideally, their outcomes would be collected as part of routine clinical surveillance without requiring specific parental consent for inclusion in the analyses.

2. Bayesian Inference. Rare conditions and subgroup analyses will continue to pose a challenge for future NRN studies. In such settings with small sample sizes, Bayesian methods can be used to estimate the probability of treatment benefit and harm even for a conventionally underpowered trial.[23] These probabilities cannot be calculated with traditional frequentist methods, and Bayesian analyses are needed to calculate them and the probability of a clinically important treatment effect.[24]

   For interventions where prior studies in the area exist, as is the case of the ongoing Late Hypothermia and Preemie Hypothermia trials, use of this prior information can potentially reduce the required sample size compared with a frequentist design. Additionally, by incorporating the current state of knowledge through the use of informative prior distributions, Bayesian analyses provide a natural way to formally incorporate empirical evidence on the plausible magnitude of treatment effects and have been proposed as a way to temper large effects that may be spurious.[25,26] Prior information that was skeptical of large treatment effects was used for the design of the NRN Late Hypothermia and Preemie

Hypothermia trials and for the analyses of the NRN Aggressive Phototherapy Trial.[27,28]

If prior evidence or even strong beliefs from investigators exist for a particular intervention, then enthusiastic or skeptical priors can be used as a sensitivity analysis to ensure that trial results are robust to differing degrees of prior belief regarding the efficacy of the intervention. The resulting final or posterior distributions for the treatment effect may differ for different levels of prior belief, particularly for small sample sizes, and investigators may reach different conclusions for a particular study. Although this subjectivity of prior belief is often offered as a main drawback of a Bayesian approach, it can also be seen as an advantage because it formalizes how experts with differing preexisting opinions will view the results. Different prior distributions and all other aspects of a Bayesian trial design and analysis plan should be evaluated with simulation studies (using different scenarios to represent a range of potential treatment effects) to ensure optimal trial operating characteristics (i.e., adequate control for type I and II errors).[29,30]

3. Other Designs. Future studies may also take advantage of emerging study designs for minimally sized studies. Group sequential designs reduce sample size by stopping studies early if effect sizes differ from what is expected, and internal pilot/adaptive designs[31] use interim power analysis to adjust the sample size for misspecification of parameters, while internal pilot with interim analysis designs combine the advantages of both.[32]

Randomization

Proper randomization, conducted carefully and with adequate thought, is essential for preserving the integrity of any RCT. Trials in neonates, however, often present special issues for consideration, some of which are discussed below.

Multiple Births. Multiple births are fairly common in the extremely premature population that is the subject of most NRN trials, with more than a quarter of babies born at 22–28 weeks gestation reported to fit this description.[1] Because outcomes of infants from the same birth may be correlated because of shared environmental and genetic factors, the statistical question then concerns whether babies from the same birth should be randomized individually or together (i.e., whether to randomize each baby or each mother/pregnancy). The advantage of the latter approach, which is akin to a cluster randomized trial (with the cluster here being the mother) is that it forces the analysis to account for any correlation in outcomes among babies from the same birth. This is also the natural approach to adopt if the randomized intervention being tested will be conducted on the mother antenatally or in the delivery room, thus ensuring that babies from the same birth are, by design, exposed to the same intervention.

For interventions conducted on babies after birth, however, cluster randomization of multiples to the same treatment arm can present substantial downsides. First, the correlation in outcomes between multiples may not be especially high because less than 10% of multiple births in the very low birth weight population are expected to be monozygotic, and thus share identical genetic makeup.[33,34] So, although monozygotic multiples may be expected to have higher correlation in outcomes, for dyzygotic multiples other factors such as gender, birth weight and sex discordance, and events during the NICU course may be far more important in determining outcomes.[35–37] Thus, cluster randomization may be trying to solve a problem that may be substantively insignificant.

Second, any small correlation in outcomes, if present among multiple births, can be accounted for in the analysis using traditional methods for correlated data analysis (such as generalized estimating equation [GEE] or random effects models) without recourse to group randomization of multiples. It is interesting to note that one of the few publications in this area did not note any advantages to specialized modeling for correlated data in the case of binary outcomes (the type of outcomes most commonly used for large multicenter neonatal trials),[33] although others have reported different results.[38–40]

Third, cluster randomization entails the use of mixed models or GEE for their analysis, regardless of whether any appreciable correlation is actually present in the outcomes of interest. Such methods, by having to estimate additional covariance parameters from the same set of data, usually produce estimates that are less precise with wider confidence intervals, compared to ordinary regression models that do not account for correlations. It is also noteworthy that in cluster randomization the unit of randomization (the cluster or the mother) is different from the unit of analysis and inference (the baby), and thus inferences have to be tailored and interpreted accordingly.

Fourth, it is often difficult to develop precise sample size and power calculations for a cluster randomized trial because of the lack of prior data on the size of the intracluster correlation for different outcomes (which would vary depending on the outcome, the composition of multiple births in the total sample, and the composition of monozygotic births among the multiple births). Typically, accounting for this correlation increases the sample size required, which has implications for both trial budgets and the amount of time needed to finish accrual.

In summary, although group randomization of multiple births may be conceptually appealing, there are a number of methodologic issues that should be carefully considered before a decision is made to use this approach. The NRN has used individual randomization for almost all its randomized trials, with group randomization only used for the SUPPORT trial, which sought antenatal consent, and where one of the two interventions took place in the delivery room.[41] We recognize that ethical issues such as parent preference

should also be strongly considered in the trial design process, although these are outside the scope of this report.[42] This remains a fertile area for further methodological research that can potentially use comprehensive simulations over a wide range of scenarios to conduct a statistical cost-benefit analysis of the relative merits of either approach to randomization in the population of extremely premature babies.

Clinic Interventions. Sometimes, interventions need to be tested at the clinic (or NICU) level because they involve broad-based unit practice or process changes, or interventions that cannot be individually delivered to patients. Even if the interventions can be theoretically tailored individually, sometimes that is not possible in reality without "contaminating" the practice across other unit clinical staff, in which case again the intervention needs to be tested at the unit level. Such interventions are tested in randomized trials using cluster randomization where the clinics or NICUs, rather than individual babies, are randomized to different treatment arms.[43] Because the unit of randomization for these trials is the center (or NICU), the power and sample size requirements and the available precision for treatment effect estimates for these trials is driven more by the total number of centers in the trial than the total number of babies enrolled.

A variation on cluster randomization involves cluster randomized crossover trials where each cluster (or clinic) receives each intervention at least once in separate periods of time.[44,45] During each time period the cluster may contain different babies, the same babies, or a mixture of both different and same babies.[46] Such designs, if practically feasible, are typically more efficient and require less sample size than traditional cluster randomized designs if the cluster environment remains similar between time periods. The stepped wedge cluster randomized trial is another novel variation on the crossover cluster design, which involves random and sequential crossover of clusters from control to intervention until all clusters are exposed. It is increasingly being used to study Phase IV–type effectiveness interventions, where efficacy has already been established and it would be ethically problematic not to offer the intervention to all clusters, but logistically challenging to start the intervention at many sites simultaneously.[47]

With the exception of the cluster randomized Benchmarking trial,[48] such trials have been rare in the NRN because of the limited number of centers available to randomize, which makes it difficult to estimate treatment effects with adequate precision. The statistical challenges outlined for cluster randomized trials in the previous discussion all apply here as well. However, cluster randomized trials remain the most rigorous means by which to study interventions instituted at the center level.

Multiple Interventions: Factorial Designs

At times, testing the efficacy of two or more treatments concurrently may help determine whether there is any advantage in using them together and use resources more efficiently by

studying two treatments in the same trial.[49] However, efficiency can only be achieved for factorial designs if one can reasonably assume that there is no statistical interaction between the different treatments being tested (i.e., the effect of one is not changed in the presence of the other). Factorial designs also provide the most rigorous means to test for the presence of statistical interactions among the treatments being studied. However, the goals of efficiency and studying interactions are in conflict, and one can only study interactions in a two-treatment (or 2×2) factorial design with adequate power if the sample size is roughly four times higher than what would be required under an assumption of no statistical interaction. To be studied together in a factorial design, certain conditions need to hold:

1. It is possible to administer the treatments/interventions together without having to modify one because of the presence of the other.
2. All possible treatment group combinations, including placebo, are ethically acceptable.
3. All treatment group combinations are of scientific interest.
4. The different treatments being tested do not have the same mechanism of action (because otherwise either one will answer the scientific question).

When the primary interest in conducting a factorial study is to test the individual effects of two treatments or interventions together simply for efficiency, and not to study their statistical interaction, an understanding of how the two treatments jointly affect the outcome of interest in comparison to how they independently affect outcomes is still essential for appropriate power and sample size calculations. Clinically, the no interaction assumption often used in a factorial design means that the treatment effect of Treatment A in a population of individuals already treated with Treatment B is identical to the effect of Treatment A when no other treatment has been implemented. Clinicians often feel that such an assumption is not tenable in that a second treatment often achieves diminished results if some benefit is achieved by application of another treatment. Failure to account for this potential diminished return (typically characterized as a sub-additive interaction effect) will result in an underpowered factorial design.

In the absence of historic information about statistical interactions between different treatments being studied, and in recognition of the prohibitive sample size required to examine such interactions with adequate statistical power, in general factorial design trials in the NRN have assumed no independence among the studied treatments, but tested for the presence of such interactions once the trials were complete. Earlier NRN trials have tested permissive ventilation and steroids in reducing BPD,[50,51] and delivery room CPAP or surfactant to reduce BPD, in conjunction with different oxygen saturation targets to reduce ROP.[10,41] More recently, the Optimizing Cooling trial tested the efficacy of longer and deeper cooling in improving neurodevelopmental outcomes of neonatal encephalopathy using a 2×2 factorial design.[52]

Statistical Interim Monitoring and Stopping Rules

Most NRN trials have a composite primary endpoint of death or morbidity, which complicates safety monitoring when trials can potentially be stopped early because of

differences in mortality at interim analyses while the composite outcome will be tested at the end of the trial.[53] These trials require interim monitoring strategies with appropriate partitioning of the overall Type I error rate. For morbidities such as NEC that may manifest over time, survival analysis can be used to model time to onset of disease, with deaths censored.[54] Typically, to ensure patient safety, we monitor adverse events more frequently (e.g., every 100 babies for the ongoing Hydrocortisone trial) than measures of efficacy (e.g., 50% and 75% of primary outcome accrual for the same trial), and use relatively liberal Pocock statistical bounds to ensure interim safety and more conservative O'Brien Fleming bounds to determine interim efficacy.

Although the Hydrocortisone trial provides an example of a typical approach to interim testing, monitoring plans are tailored to each individual study. For example, in the ongoing Preemie Hypothermia trial designed using Bayesian principles, interim safety is monitored using posterior probability of treatment harm (based on a neutral prior probability of treatment benefit), while interim efficacy and futility are conservatively monitored using the posterior probability of treatment benefit (based on a skeptical prior probability of treatment benefit for efficacy monitoring, and an enthusiastic prior probability of treatment benefit for futility monitoring).

Bayesian monitoring approaches have potential advantages[29,55,56] that include incorporation of results from prior trials to better assess the likelihood of treatment benefit or harm and ensure that treatment recommendations are well justified based on all relevant trials.[57] For monitoring of trials, at a given interim analysis the posterior probability of the treatment effect is computed from the prior probability and the interim data. Stopping guidelines need to be prespecified for each type of interim monitoring. When monitoring a trial for efficacy, a large posterior probability threshold (i.e., > 97%) would be required to stop a trial early for benefit. Conversely, when considering stopping for futility, a stopping guideline would require that the probability of treatment benefit is very low (i.e., < 10%). The exact probability thresholds will differ for different interventions and patient populations. For example, a particularly high probability of a clinically meaningful benefit might be required for therapies that are invasive, hazardous, or extremely expensive. Thus, the appropriate stopping probability thresholds need to be discussed between the statistician and clinical investigators for each study.

A Bayesian approach can also incorporate a wide range of viewpoints and indicate the magnitude of the difference between treatment groups that would be needed at the end of the trial to convince those who are skeptical and those who were enthusiastic about the value of the therapy prior to the trial.[24,57] For example, to stop early for benefit a Bayesian analysis would use a prior skeptical of treatment benefit. The resulting posterior probability should be sufficient to convince an investigator who was skeptical of any treatment benefit at the beginning of the trial that the therapy is beneficial. To stop for futility, we would take the position of an enthusiastic investigator with strong prior belief in treatment benefit and assess whether there is sufficiently convincing evidence that there is little chance of benefit from the intervention.

When deciding whether to continue recruitment, pause, or terminate a trial, a Data and Safety Monitoring Committee (DSMC) can then not only weigh the current evidence for benefit, harm, or futility from the posterior probability but also formally assess how these results would be viewed by clinicians with different levels of skepticism or enthusiasm. A DSMC can use Bayesian interim analyses to evaluate the "totality of available evidence" and consider whether the results of a trial would be convincing to skeptical, enthusiastic, and neutral clinical investigators. Decisions to stop a trial early based on the best available data from all relevant trials and on the identification of clinically meaningful differences are most likely to be convincing to patients, clinicians, and investigators.

Center Effects

Although the diversity of NRN NICUs increases the external validity of its studies, large center differences may also obscure effects of interest. Thus, NRN trials are stratified by center whenever possible, with adjustment for center effects in analyses.[58,59] When the sample size is too small to account for centers as fixed effects, or when center-independent predictions are desired,[60] generalized linear mixed models (GLMM) can include center as a random effect, accounting for both within- and between-center variations.

Hierarchical models such as GLMM have the added benefit of being able to incorporate appropriate center-level data on clinical practice, if available, to help explain and quantify (and not merely account for) center differences.[61]

Thus, recently, to evaluate whether differences in initiation of active treatment for extremely premature babies across hospitals may explain the wide center differences in mortality outcomes across the NRN, we used multivariable multilevel logistic regression models to assess clustering of active treatment at the hospital level, by gestational age at birth, after accounting for differences in patient characteristics. Models included infant-level receipt of active treatment as a binary outcome and were adjusted for infant-level characteristics known at birth, and the overall active treatment rate at the hospital level. These models were used to calculate the intra-class correlation for active treatment (i.e., the proportion of variation in active treatment that was attributable to an infant's hospital of birth) by gestational age at birth.[62]

Causal Inference from Observational Studies

Clinical trials are the gold standard for developing the evidence base of neonatal medicine. However, randomized trials are often not possible because of ethical considerations, lack of equipoise, or lack of resources. For example, the question of benefit from antenatal steroids for babies born at the edges of viability (25 weeks and under) has not been adequately addressed by a randomized trial, but the ubiquitous use of this treatment in all extremely premature babies does not make a trial viable (ethically or in terms of equipoise) anymore.[63] Thus, many open questions in neonatal care can only be addressed by carefully conducted observational studies. Some of the challenges in using observational studies to develop the evidence base for neonatal medicine are briefly outlined below:

Covariates: Selection and control for proper covariates and confounders is essential to proper inference in an observational study. Prospectively planned observational studies in the NRN

strive to develop an a priori covariate selection strategy to avoid problems of reverse causation, circular reasoning, and spurious results. The elements of such a strategy typically include the following steps:

- **a.** Principal covariates and confounders are identified from the literature.
- **b.** A primary tier of covariates is identified (typically center, gestational age, gender, exposure to antenatal steroids, and SGA or birth weight) that are always adjusted for, regardless of statistical or clinical significance (so that the effect estimates of interest can be reported as adjusted for such covariates).
- **c.** A secondary tier of covariates is identified that can be adjusted for if indicated by the data or extant literature.
- **d.** Depending on the goals of the analysis (descriptive, hypothesis testing, or prediction), rules for final covariate selection (from the second tier) are set forth a priori based on both statistical and clinical significance.

One of the frequent difficulties faced in neonatal studies is the lack of a universally accepted and comprehensive index measure of baseline level of illness severity. The lack of such a measure can lead to foregone conclusions whereby babies that are sicker at the outset are exposed to more clinical interventions and often have worse outcomes. Different NRN studies have used various measures of baseline level of sickness such as baseline probability of death or NDI (the ongoing NEST trial) or receiving mechanical ventilation for the first 7 days of life.[64]

Causal Inference: Causal inference, whereby we can impart causality and not just association between risk factors (or treatments/interventions) and outcomes from observational studies, is increasingly being used in situations where randomized trials are not possible. Typically, this involves carefully formulating the causal question to be answered, stating the underlying assumptions, using novel analysis methods, and finally using sensitivity analyses to explore the robustness of the conclusions. In the NRN we have used propensity scores modeling[65,66] to compare outcomes for nonrandomized treatments where we use a two-step modeling procedure to reduce bias in such comparisons. For example, in a recent study to examine associations between inhaled nitric oxide (iNO) use and severe BPD or death, we used this approach to even out any imbalance between the iNO treated and untreated groups with regard to baseline risk factors for iNO treatment (to the extent permitted by available data), because iNO treatment was not randomized and there were no other study design features that might have checked such imbalances when iNO treatment was initiated.[67,68]

A propensity score was created from a logistic regression model quantifying the probability of receiving iNO treatment, as predicted by a set of well-chosen covariates in the propensity model. Once we obtained the propensity scores, they were ranked and divided into similar groups based on their ranks. Thus, 12 strata were created, each consisting of both iNO treated and control infants that were balanced on key covariates for these two types of infants. A variable representing these strata was included as a predictor in a second logistic

regression analysis model to compare the primary outcome (severe BPD or death) between the two treatment groups.

Propensity scores modeling offers a useful conceptual framework to formulate a causal inference question from observational data. In addition, the two-step modeling approach can sometimes circumvent model-fitting and precision problems faced by ordinary regression. However, it is useful to remember that both methods are still vulnerable to the problem of unmeasured or unknown covariates that may be imbalanced across the comparison groups, and can thus produce biased inference.

Longitudinal Analyses

Increased survival of babies at high risk of long-term morbidities emphasizes the need for long-term follow-up. Neurodevelopmental deficiencies (such as IQ, behavioral problems, or conditions such as autism spectrum disorders or attention deficit disorder) may not manifest until later childhood and can only be understood in the context of a child's overall developmental trajectory. Longitudinal analyses are therefore essential because they are more powerful than cross-sectional studies in terms of explanatory power and statistical efficiency.[69] For example, the Hypothermia extended follow-up study compared the neurodevelopmental trajectories of treated and control children.[70] The downside to a longitudinal study are usually higher costs, more attrition and missing data over time, and greater potential for informative missing data at later time points to lead to biased inference.

Multiple Endpoints

In some situations, multiple outcomes may measure different aspects of the same underlying syndrome. We have developed a multivariate approach to model such outcomes simultaneously. This approach conserves degrees of freedom, adjusts for correlations among the outcomes, and avoids multiple comparisons, while allowing for the estimation of outcome-specific effects.[71,72]

Biomarkers

Because neonatal research is trending toward longer follow-up to school age, we do not foresee NRN trials using biomarkers as surrogate primary outcomes. However, we do anticipate interest in the validation of prognostic and predictive biomarkers for long-term outcomes for conducting early safety monitoring, counseling parents, identifying candidates for early intervention, and designing studies (as eligibility criteria or stratification factors). Technologies such as MRI and aEEG evaluated in children followed to school age (Hypothermia and SUPPORT trials) and the NRN Biorepository for collection of omics data provide unique opportunities for biomarker identification in extremely premature babies. Biomarker validation methods continue to evolve, as do relevant guidelines such as multistep biomarker development process recommendations by the MicroArray Quality Control-II,[73] ACCE evaluation criteria (Analytic and Clinical validity, Clinical utility, Ethical, legal, and social issues),[74] and recommendations from the Institute of Medicine.[75] Statistical geneticists in the NRN have contributed to advancements in the field both as coauthors of these guidance documents[73] and as developers of innovative methods for the analysis of omics data as illustrated below.

Genetics and Omics Data

The NRN has conducted candidate gene and genome-wide association (GWA) studies of BPD, ROP, ICH, and NEC in extremely premature babies using innovative approaches such as gene/burden-based testing to minimize multiple testing issues and pathways analysis to help interpret the data.[76–78] We have also developed new methods for incorporating covariates into GWA studies with maximal power while minimizing biases.[77] As whole genome sequencing becomes less expensive it will likely replace GWA studies, with an associated increase in bioinformatics burden.[79] Future NRN research may also include studies of the microbiome, metagenomics, metabolomics, and proteomics. For example, metagenomics may explore biomarkers for NEC and brain injury in extremely premature and encephalopathic neonates, respectively.

Pharmacokinetic Studies

Complicating factors in PK studies in extremely premature infants include sparse blood samples over time because of infant size, switching of administration mode from IV to oral as an infant matures, and endogenous synthesis of study compound by the infants. We faced each of these three challenges in the NRN Phase II RCT for safety and PK dose-determination of multiple doses of myo-inositol administered every 12 hours for several days (preparatory to a Phase III RCT to prevent ROP). Our solution was to extend and combine population-PK one- and two-compartment models for IV and oral administration into a single model that accounted for change in administration mode from IV to oral and measuring endogenous synthesis of inositol. In addition, our model efficiently combined the sparse repeated measures data from each infant using population-PK methods based on nonlinear mixed effects models.[80]

CONCLUSION

We have discussed some of the specific methodological challenges in the statistical design and analysis of RCTs and observational studies in extremely premature neonates. Although some of these challenges (such as rare outcomes) may not be unique to neonatology, the aggregate constellation of patient population features discussed above is fairly unique and calls for broad understanding and careful consideration of these issues during the conception phase for any rigorous study in this population. As described above, the NRN has developed pragmatic solutions for many of these situations. However, there continue to be unanswered questions and ample room for substantial advances in methodology and new ways of thinking to further refine the design and analysis of neonatal trials and studies and continue to build the evidence base for neonatal medicine.

REFERENCES

1. Stoll BJ, Hansen NI, Bell EF, et al. Trends in Care Practices, Morbidity, and Mortality of Extremely Preterm Neonates, 1993–2012. JAMA. 2015; 314(10):1039–1051. [PubMed: 26348753]
2. Costeloe K, Hennessy E, Myles J, Draper E. EPICure 2: survival and early morbidity of extremely preterm babies in England: changes since. [Accessed March 8, 2016] E-PAS2008:5365.1. Available at http://www.abstracts2view.com/pasall.
3. Fischer N, Steurer MA, Adams M, Berger TM, Swiss Neonatal N. Survival rates of extremely preterm infants (gestational age <26 weeks) in Switzerland: impact of the Swiss guidelines for the

care of infants born at the limit of viability. Arch Dis Child Fetal Neonatal Ed. 2009; 94(6):F407–F413. [PubMed: 19357122]

4. World Health Organization, March of Dimes, The Partnership for Maternal, Newborn, & Child Health, Save the Children. Geneva, Switzerland: World Health Organization; 2012. Born too soon: the global action report on preterm birth. http://www.who.int/pmnch/media/news/2012/201204_borntoosoon-report.pdf [Accessed March 8, 2016]
5. Hintz SR, Kendrick DE, Wilson-Costello DE, et al. Early-childhood neurodevelopmental outcomes are not improving for infants born at <25 weeks' gestational age. Pediatrics. 2011; 127(1):62–70. [PubMed: 21187312]
6. Hack M, Wilson-Costello DE, Friedman H, Minich N, Siner B. Early childhood outcomes of infants born at the limits of viability have not improved in 2000–2004. [Accessed March 8, 2016] E-PAS2008:5365.21. Available at http://www.abstracts2view.com/pasall.
7. Colin AA, McEvoy C, Castile RG. Respiratory morbidity and lung function in preterm infants of 32 to 36 weeks' gestational age. Pediatrics. 2010; 126(1):115–128. [PubMed: 20530073]
8. Raju TN, Higgins RD, Stark AR, Leveno KJ. Optimizing care and outcome for late-preterm (near-term) infants: a summary of the workshop sponsored by the National Institute of Child Health and Human Development. Pediatrics. 2006; 118(3):1207–1214. [PubMed: 16951017]
9. Tyson JE, Walsh M, D'Angio CT. Comparative effectiveness trials: generic misassumptions underlying the SUPPORT controversy. Pediatrics. 2014; 134(4):651–654. [PubMed: 25201795]
10. Carlo WA, Finer NN, Walsh MC, et al. Target ranges of oxygen saturation in extremely preterm infants. N Engl J Med. 2010; 362(21):1959–1969. [PubMed: 20472937]
11. [Accessed March 8, 2016] ClinicalTrials.gov website. Transfusion of Prematures Trial (TOP). 2015 Jul 21. Available at https://clinicaltrials.gov/ct2/show/NCT01702805?term=transfusion+of+prematures&rank=1
12. Sinclair JC. Weighing risks and benefits in treating the individual patient. Clin Perinatol. 2003; 30(2):251–268. [PubMed: 12875353]
13. Walter SD, Sinclair JC. Uncertainty in the minimum event risk to justify treatment was evaluated. J Clin Epidemiol. 2009; 62(8):816–824. [PubMed: 19216053]
14. Marcucci M, Sinclair JC. A generalised model for individualising a treatment recommendation based on group-level evidence from randomised clinical trials. BMJ Open. 2013; 3(8):pii e003143.
15. [Accessed March 8, 2016] ClinicalTrials.gov website. Hydrocortisone for BPD. 2015 Jul 21. Available at https://clinicaltrials.gov/ct2/show/NCT01353313?term=hydrocortisone+for+BPD&rank=1
16. Nolen TL, Hudgens MG. Randomization-Based Inference within Principal Strata. J Am Stat Assoc. 2011; 106(494):581–593. [PubMed: 21987597]
17. Chen YH. Maximum likelihood analysis of semicompeting risks data with semiparametric regression models. Lifetime Data Anal. 2012; 18(1):36–57. [PubMed: 21850528]
18. Xu J, Kalbfleisch JD, Tai B. Statistical analysis of illness-death processes and semicompeting risks data. Biometrics. 2010; 66(3):716–725. [PubMed: 19912171]
19. Andersen PK, Pohar Perme M. Inference for outcome probabilities in multi-state models. Lifetime Data Anal. 2008; 14(4):405–431. [PubMed: 18791824]
20. King M, Nazareth I, Lampe F, et al. Impact of participant and physician intervention preferences on randomized trials: a systematic review. JAMA. 2005; 293(9):1089–1099. [PubMed: 15741531]
21. Gelman, A.; Meng, X. Applied Bayesian modeling and causal inference from incomplete data perspectives. New York, NY: John Wiley; 2004.
22. Rubin DB. The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. Stat Med. 2007; 26(1):20–36. [PubMed: 17072897]
23. Lilford RJ, Thornton JG, Braunholtz D. Clinical trials and rare diseases: a way out of a conundrum. BMJ. 1995; 311(7020):1621–1625. [PubMed: 8555809]
24. Spiegelhalter, DJ.; Abrams, KR.; Myles, JP. Bayesian approaches to clinical trials and health-care evaluation. Chichester, UK: John Wiley & Sons; 2004.
25. Greenland S. Putting background information about relative risks into conjugate prior distributions. Biometrics. 2001; 57(3):663–670. [PubMed: 11550913]

26. Pedroza C, Han W, Truong VT, Green C, Tyson JE. Performance of informative priors skeptical of large treatment effects in clinical trials: A simulation study. Stat Methods Med Res. 2015
27. Morris BH, Oh W, Tyson JE, et al. Aggressive vs. conservative phototherapy for infants with extremely low birth weight. N Engl J Med. 2008; 359(18):1885–1896. [PubMed: 18971491]
28. Tyson JE, Pedroza C, Langer J, et al. Does aggressive phototherapy increase mortality while decreasing profound impairment among the smallest and sickest newborns? J Perinatol. 2012; 32(9):677–684. [PubMed: 22652561]
29. Berry DA. Bayesian clinical trials. Nat Rev Drug Discov. 2006; 5(1):27–36. [PubMed: 16485344]
30. Food and Drug Administration. Guidance for the use of Bayesian statistics in medical device clinical trials. Rockville, MD: Center for Biologics Evaluation and Research (CBER), Office of Communication, Outreach and Development; 2010 Feb 5.
31. Coffey CS, Kairalla JA. Adaptive clinical trials: progress and challenges. Drugs R D. 2008; 9(4): 229–242. [PubMed: 18588354]
32. Kairalla JA, Muller KE, Coffey CS. Combining an Internal Pilot with an Interim Analysis for Single Degree of Freedom Tests. Commun Stat Theory Methods. 2010; 39(20):3717–3738. [PubMed: 21037942]
33. Shaffer ML, Kunselman AR, Watterberg KL. Analysis of neonatal clinical trials with twin births. BMC Med Res Methodol. 2009; 9:12. [PubMed: 19245713]
34. Luke B. What is the influence of maternal weight gain on the fetal growth of twins? Clin Obstet Gynecol. 1998; 41(1):56–64. [PubMed: 9504224]
35. Ananth CV, Demissie K, Hanley ML. Birth weight discordancy and adverse perinatal outcomes among twin gestations in the United States: the effect of placental abruption. Am J Obstet Gynecol. 2003; 188(4):954–960. [PubMed: 12712093]
36. Marttila R, Kaprio J, Hallman M. Respiratory distress syndrome in twin infants compared with singletons. Am J Obstet Gynecol. 2004; 191(1):271–276. [PubMed: 15295378]
37. Vergani P, Locatelli A, Ratti M, et al. Preterm twins: what threshold of birth weight discordance heralds major adverse neonatal outcome? Am J Obstet Gynecol. 2004; 191(4):1441–1445. [PubMed: 15507980]
38. Yelland LN, Sullivan TR, Makrides M. Accounting for multiple births in randomised trials: a systematic review. Arch Dis Child Fetal Neonatal Ed. 2015; 100(2):F116–F120. [PubMed: 25389142]
39. Hibbs AM, Black D, Palermo L, et al. Accounting for multiple births in neonatal and perinatal trials: systematic review and case study. J Pediatr. 2010; 156(2):202–208. [PubMed: 19969305]
40. Yelland LN, Sullivan TR, Pavlou M, Seaman SR. Analysis of Randomised Trials Including Multiple Births When Birth Size Is Informative. Paediatr Perinat Epidemiol. 2015; 29(6):567–575. [PubMed: 26332368]
41. Finer NN, Carlo WA, Walsh MC, et al. Early CPAP versus surfactant in extremely preterm infants. N Engl J Med. 2010; 362(21):1970–1979. [PubMed: 20472939]
42. Bernardo J, Nowacki A, Martin R, Fanaroff JM, Hibbs AM. Multiples and parents of multiples prefer same arm randomization of siblings in neonatal trials. J Perinatol. 2015; 35(3):208–213. [PubMed: 25341196]
43. Murray DM, Varnell SP, Blitstein JL. Design and analysis of group-randomized trials: a review of recent methodological developments. Am J Public Health. 2004; 94(3):423–432. [PubMed: 14998806]
44. Turner RM, White IR, Croudace T. Group PIPS. Analysis of cluster randomized cross-over trial data: a comparison of methods. Stat Med. 2007; 26(2):274–289. [PubMed: 16538700]
45. Parienti JJ, Kuss O. Cluster-crossover design: a method for limiting clusters level effect in community-intervention studies. Contemp Clin Trials. 2007; 28(3):316–323. [PubMed: 17110172]
46. Rietbergen C, Moerbeek M. The design of cluster randomized crossover trials. J Educ Behav Stat. 2011; 36(4):472–490.
47. Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. BMJ. 2015; 350:h391. [PubMed: 25662947]

48. Walsh M, Laptook A, Kazzi SN, et al. 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. Pediatrics. 2007; 119(5):876–890. [PubMed: 17473087]
49. Piantadosi, S. Clinical trials: a methodological perspective. New York, NY: John Wiley; 2010.
50. Stark AR, Carlo WA, Tyson JE, et al. Adverse effects of early dexamethasone in extremely-low-birth-weight infants. National Institute of Child Health and Human Development Neonatal Research Network. N Engl J Med. 2001; 344(2):95–101. [PubMed: 11150359]
51. Carlo WA, Stark AR, Wright LL, et al. Minimal ventilation to prevent bronchopulmonary dysplasia in extremely-low-birth-weight infants. J Pediatr. 2002; 141(3):370–374. [PubMed: 12219057]
52. Shankaran S, Laptook AR, Pappas A, et al. Effect of depth and duration of cooling on deaths in the NICU among neonates with hypoxic ischemic encephalopathy: a randomized clinical trial. JAMA. 2014; 312(24):2629–2639. [PubMed: 25536254]
53. Chen YH, DeMets DL, Lan KK. Monitoring mortality at interim analyses while testing a composite endpoint at the final analysis. Control Clin Trials. 2003; 24(1):16–27. [PubMed: 12559639]
54. Meinzen-Derr J, Poindexter B, Wrage L, Morrow AL, Stoll B, Donovan EF. Role of human milk in extremely low birth weight infants' risk of necrotizing enterocolitis or death. J Perinatol. 2009; 29(1):57–62. [PubMed: 18716628]
55. Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR. Methods in health service research. An introduction to bayesian methods in health technology assessment. BMJ. 1999; 319(7208):508–512. [PubMed: 10454409]
56. Harrell FE Jr, Shih YC. Using full probability models to compute probabilities of actual interest to decision makers. Int J Technol Assess Health Care. 2001; 17(1):17–26. [PubMed: 11329842]
57. Fayers PM, Ashby D, Parmar MK. Tutorial in biostatistics Bayesian data monitoring in clinical trials. Stat Med. 1997; 16(12):1413–1430. [PubMed: 9232762]
58. Cotten CM, Oh W, McDonald S, et al. Prolonged hospital stay for extremely premature infants: risk factors, center differences, and the impact of mortality on selecting a best-performing center. J Perinatol. 2005; 25(10):650–655. [PubMed: 16079906]
59. Vohr BR, Wright LL, Dusick AM, et al. Center differences and outcomes of extremely low birth weight infants. Pediatrics. 2004; 113(4):781–789. [PubMed: 15060228]
60. Tyson JE, Parikh NA, Langer J, et al. Intensive care for extreme prematurity--moving beyond gestational age. N Engl J Med. 2008; 358(16):1672–1681. [PubMed: 18420500]
61. Woodward A, Das A, Raskin IE, Morgan-Lopez AA. An exploratory analysis of treatment completion and client and organizational factors using hierarchical linear modeling. Eval Program Plann. 2006; 29(4):335–351. [PubMed: 17950862]
62. Rysavy MA, Li L, Bell EF, et al. Between-hospital variation in treatment and outcomes in extremely preterm infants. N Engl J Med. 2015; 372(19):1801–1811. [PubMed: 25946279]
63. Carlo WA, McDonald SA, Fanaroff AA, et al. Association of antenatal corticosteroids with mortality and neurodevelopmental outcomes among infants born at 22 to 25 weeks' gestation. JAMA. 2011; 306(21):2348–2358. [PubMed: 22147379]
64. Ehrenkranz RA, Das A, Wrage LA, et al. Early nutrition mediates the influence of severity of illness on extremely LBW infants. Pediatr Res. 2011; 69(6):522–529. [PubMed: 21378596]
65. Morriss FH Jr, Saha S, Bell EF, et al. Surgery and neurodevelopmental outcome of very low-birth-weight infants. JAMA Pediatr. 2014; 168(8):746–754. [PubMed: 24934607]
66. Truog WE, Nelin LD, Das A, et al. Inhaled nitric oxide usage in preterm infants in the NICHD Neonatal Research Network: inter-site variation and propensity evaluation. J Perinatol. 2014; 34(11):842–846. [PubMed: 24901452]
67. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983; 70(1):41–55.
68. D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Stat Med. 1998; 17(19):2265–2281. [PubMed: 9802183]
69. Diggle, P.; Heagerty, P.; Liang, K.; Zeger, S. Analysis of longitudinal data. London, UK: Oxford University Press; 2002.

70. Shankaran S, Pappas A, McDonald SA, et al. Childhood outcomes after hypothermia for neonatal encephalopathy. N Engl J Med. 2012; 366(22):2085–2092. [PubMed: 22646631]

71. Bada HS, Bauer CR, Shankaran S, et al. Central and autonomic system signs with in utero drug exposure. Arch Dis Child Fetal Neonatal Ed. 2002; 87(2):F106–F112. [PubMed: 12193516]

72. Das A, Poole WK, Bada HS. A repeated measures approach for simultaneous modeling of multiple neurobehavioral outcomes in newborns exposed to cocaine in utero. Am J Epidemiol. 2004; 159(9):891–899. [PubMed: 15105182]

73. Shi L, Campbell G, Jones WD, et al. The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models. Nat Biotechnol. 2010; 28(8):827–838. [PubMed: 20676074]

74. Haddow, J.; Palomaki, G. ACCE: a model process for evaluating data on emerging genetic tests. In: Khoury, M.; Little, J.; Burke, W., editors. Human genome epidemiology: a scientific foundation for using genetic information to improve health and prevent disease. New York, NY: Oxford University Press; 2003. p. 217-233.

75. Institute of Medicine. Evolution of translational omics: lessons learned and the path forward. 2012 Mar 23. [Accessed March 8, 2016] Available at http://www.iom.edu/Reports/2012/Evolution-of-Translational-Omics.aspx.

76. Cooley PC, Clark R, Folsom R, Page G. Genetic inheritance and genome wide association stastitical test performance. J Proteomics Bioinform. 2010; 3(12):321–325.

77. Page, G.; Garge, N.; Johnson, EO. To adjust or not to adjust, how and when to incorporate covariates into GWA studies. In: Page, G.; Garge, N.; Johnson, EO., editors. Proceedings of the American Society of Human Genetics. San Francisco, CA: Cell Press; 2012.

78. Deelen J, Uh HW, Monajemi R, et al. Gene set analysis of GWAS data for human longevity highlights the relevance of the insulin/IGF-1 signaling and telomere maintenance pathways. Age (Dordr). 2013; 35(1):235–249. [PubMed: 22113349]

79. Platts AE, Land SJ, Chen L, et al. Massively parallel resequencing of the isogenic Drosophila melanogaster strain w(1118); iso-2; iso-3 identifies hotspots for mutations in sensory perception genes. Fly (Austin). 2009; 3(3):192–203. [PubMed: 19690466]

80. Phelps DL, Ward RM, Williams RL, et al. Pharmacokinetics and safety of a single intravenous dose of myo-inositol in preterm infants of 23–29 wk. Pediatr Res. 2013; 74(6):721–729. [PubMed: 24067395]

Table 1

Selection of Major NRN Studies Conducted in the Last 10 Years

Study Title*	Statistical Design Features	Statistical Analysis Features
Active Randomized Clinical Trials		
Hydrocortisone for BPD	Powered for joint evaluation of efficacy conditional on safety	Two-stage sequential evaluation of efficacy conditional on safety
Late Hypothermia	RCT stratified by age of enrollment and level of encephalopathy	Bayesian interim monitoring for safety, futility, and efficacy Bayesian analysis of final outcomes for treatment benefit
MILK Trial	RCT stratified by birth weight (BW), center and feeding group eligibility Simulations for operating characteristics	Linear regression to estimate the adjusted mean difference in cognitive scaled scores between the two treatment groups
NEC Surgery Trial (NEST)	Comprehensive cohort design Randomization stratified by predicted baseline risk	Meta-analysis of randomized and preference cohorts after propensity modeling/adjustment in the latter
Optimizing Cooling	2×2 factorial design with stratified randomization	Marginal and within-table adjusted analyses
Transfusion of Prematures (TOP)	Pragmatic RCT with stratified randomization and pilot study of blood bank variations	Survival analyses of death Longitudinal modeling of growth Economic evaluation
Hydrocortisone for Cardiovascular Insufficiency	Randomized, multicenter, double-blind, placebo-controlled trial Patients are enrolled and randomized in a variable block design, 1:1 ratio of hydrocortisone or placebo	Planned sample size reanalysis after 1 year of enrollment based on overall outcome rate
Preemie Hypothermia	RCT stratified by age of enrollment and level of encephalopathy	Bayesian interim monitoring for safety, futility, and efficacy Bayesian analysis of final outcomes for treatment benefit
Incubator Weaning	RCT stratified by center and gestational age	Median regression adjusting for the trial stratification factors
Inositol RCT	RCT stratified by center and gestational age	Separate analyses plans for publication and FDA submission
Completed Randomized Clinical Trials		
Early BP Pilot	Time-limited feasibility pilot study with 2×2 factorial design	Feasibility analyses
Vitamin E Pilot	Pilot, double-blind placebo-controlled RCT	Analyses of Vitamin E levels over time
SUPPORT	2×2 factorial design, randomization stratified by site and GA groups	Marginal outcomes analyses Interim group sequential monitoring in factorial setup Familial clustering
Inositol Multidose Trial	Double-blind phase II trial conducted under IND from FDA in three treatment groups stratified by GA	PK modeling allowing for endogenous Inositol synthesis and change from IV to oral administration during the intervention Used to establish dosage for pivotal phase III trial
Phototherapy	RCT stratified by site and BW groups	Primary outcome at 18 months (used surrogate markers for interim safety/efficacy monitoring)

Study Title*	Statistical Design Features	Statistical Analysis Features
Registry Studies		
Generic Data Base (GDB)	Registry of all NRN very low BW births	Varied analytic approaches, including multiple regression, Classification and Regression Trees, longitudinal analyses, multilevel hierarchical generalized linear models, prediction modeling, and genetic epidemiology analyses
22–26 Month Follow-Up	Registry of all NRN extremely low BW outcomes at 18 months	
Moderate Preterm Registry	Registry of moderate preterm births	
Term Control Registry		
Active Observational Studies		
ALPS aEEG	Prospective, cohort study	
Early Onset Sepsis II	Surveillance study	Incidence/prevalence rates
Completed Observational Studies		
Term Hypotension	Observational, time-limited study	Linear and logistic regression
Early Blood Pressure	Observational, time-limited study	BP profiles analyses
Early Onset Sepsis	Surveillance study	Incidence/prevalence rates
Preemie aEEG	Prospective cohort study Sample size based on false-negative rate	Interobserver reliability modeling Prediction modeling
Candida	Prospective cohort study	Likelihood ratio–based risk stratification, latent class analyses for diagnostic tests
PCV-7	Vaccine efficacy study	Multiple linear/logistic regression for geometric mean titers