Accounting for Multiple Births in Neonatal and Perinatal Trials: Systematic Review and Case Study

Anna Maria Hibbs, MD, MSCE[1],[*], **Dennis Black, PhD**[2], **Lisa Palermo, MA**[2], **Avital Cnaan, PhD**[3], **Xianqun Luan, MS**[4], **William E Truog, MD**, **Michele C Walsh, MD, MS**[1], and **Roberta A Ballard, MD.**[6]

[1] Department of Pediatrics, Case Western Reserve University and Rainbow Babies and Children's Hospital, Cleveland, OH

[2] Department of Biostatistics, University of California San Francisco, San Francisco, CA

[3] Center for Clinical and Community Research, Children's National Medical Center, Washington, D.C.

[4] Quality Data Services Inc., King of Prussia, PA

[5] Department of Pediatrics, Children's Mercy Hospitals and Clinics, Kansas City, MO

[6] Department of Pediatrics, University of California San Francisco, San Francisco, CA

Abstract

Objectives—To determine the prevalence in the neonatal literature of statistical approaches accounting for the unique clustering patterns of multiple births. To explore the sensitivity of an actual trial to several analytic approaches to multiples.

Methods—A systematic review of recent perinatal trials assessed the prevalence of studies accounting for clustering of multiples. The NO CLD trial served as a case study of the sensitivity of the outcome to several statistical strategies. We calculated odds ratios using non-clustered (logistic regression) and clustered (generalized estimating equations, multiple outputation) analyses.

Results—In the systematic review, most studies did not describe the randomization of twins and did not account for clustering. Of those studies that did, exclusion of multiples and generalized estimating equations were the most common strategies. The NO CLD study included 84 infants with a sibling enrolled in the study. Multiples were more likely than singletons to be white and were born to older mothers (p<0.01). Analyses that accounted for clustering were statistically significant; analyses assuming independence were not.

Conclusions—The statistical approach to multiples can influence the odds ratio and width of confidence intervals, thereby affecting the interpretation of a study outcome. A minority of perinatal studies address this issue.

Keywords

multiple birth; twins; clustered data; generalized estimating equations; multiple outputation; inhaled nitric oxide; preterm

Introduction

The statistical methods most commonly used to analyze the outcomes of perinatal clinical trials assume the statistical independence of each measured outcome between individuals. In other words, they assume that individuals are not correlated. However, the outcomes of twins and infants from higher order multiple births are not independent, violating the assumptions of these analytic methods. Siblings from the same gestation share genes, as well as prenatal and postnatal environmental and iatrogenic exposures. Further correlation may be induced in a randomized clinical trial if, as often happens, multiples receive the same intervention, either as a result of randomizing multiples as a cluster or equivalently, randomizing the mother if treating during pregnancy.

Multiple births are common in the neonatal intensive care unit, particularly among patients born preterm. The prevalence of infants born from multiple gestations among very low birth weight infants at NICHD Neonatal Research Network sites in 2006 ranged from 21–30% [personal communication, MCW]. Among multiples, 87% were twins, 12% were triplets, and 2% were higher-order multiples.

Analytic strategies for clustered data are standard in other medical research situations.[42] For instance, clustered analyses are routinely used when the outcomes of both eyes are measured in ophthalmology trials, for repeated longitudinal measures of the same patient, or when subjects have been cluster randomized by site.[43–56] In these cases, the vast majority of outcomes are part of a cluster group, whether that group consists of the data from an individual site or an individual patient. In such cases, using statistical tests that assume independence for clustered data has the potential to impact both the accuracy and precision of the results. This problem has also been demonstrated in cohorts that are exclusively comprised of twins.[44,57] However, unlike studies comprised entirely of twins, in most neonatal studies singletons are the majority although multiples may make up a large minority. This makes the clustering pattern in neonatal studies unique, because most of the outcomes measured are not from a cluster group, but many are. Gates et al. also demonstrated that the analytic approach to non-independence in such situations may influence the estimates of odds ratios and confidence intervals in such populations.[58] Furthermore, multiples may be different than singletons with regards to key prognostic demographic characteristics such as race and socioeconomic status, thereby increasing the importance of appropriately accounting for their correlated status and other covariates in the analysis. Therefore, neonatal populations represent a unique situation for which a standard analytic approach has not been established.

We conducted a systematic review of recent multi-center randomized clinical trials studying preterm populations to assess the prevalence of clustered analytic techniques in the neonatal and perinatal literature. We hypothesized that inappropriately weighting the correlated outcomes of multiples by not using analyses that account for their non-independence could bias the calculated odds ratios and confidence intervals, potentially altering the interpretation of trial results. As a case study to explore the sensitivity of an actual neonatal trial to the analytic approach to multiples, we used the NO CLD trial that cluster-randomized very low birth weight infants by mother to either inhaled nitric oxide or placebo.[13,59]

Methods

Systematic Review

We conducted a systematic review to obtain an overview of the way in which multiple births are handled in the neonatal and perinatal literature. Because the incidence of multiple births is higher in preterm populations, we focused the search on premature infants. In addition, we narrowed the search to multi-center randomized clinical trials, both for feasibility and because large multi-center trials require extensive collaboration between trialists and statisticians. The search was conducted on August 29, 2008 using PubMed, with the terms "preterm and multicenter" and the limits "published in the last 5 years, Humans, Randomized Controlled Trial, English, Newborn: birth-1 month." Articles were included if they were either papers describing the methods of a multicenter trial or reporting the results of the primary outcome of the trial. A primary outcome was identified if the authors directly identified it as such or if that outcome was used to determine sample size. Trials where the primary outcome was measured in the mother but not the infant were excluded. Trials where an outcome could be equally attributed to the mother or the infant, such as breast-feeding success, were included.

The NO CLD Trial

The NO CLD trial of inhaled nitric oxide was a multi-center, blinded, placebo controlled study of the impact of inhaled nitric oxide on the development of bronchopulmonary dysplasia in 582 very low birth weight premature infants.[13,59] The primary outcome was survival without bronchopulmonary dysplasia.[60,61] The study enrolled singletons and infants born from multiple gestations. Some infants were born from a multiple gestation but did not have any siblings enrolled in the trial if those siblings did not meet eligibility criteria, were deceased, or consent was not given by the parents.

Due to expressed parental preference in previous trials, siblings from the same gestation were randomized to the same treatment. Given the cluster-randomized design, analyses accounting for clustering were planned a priori, and sample size calculations were based on clustered analyses. Generalized estimating equations were proposed as an analysis method in the original protocol, as well as random selection of a sibling from each gestation. Generalized estimating equations are an extension of generalized linear models and a form of regression that accounts for variance in clustered data.(10, 14) Multiple outputation, or within-cluster resampling, samples independent, uncorrelated data sets from the original data set, analyzes it, and combines the repeated resampling results.[48,62] This method was published during enrollment for the NO CLD study and was recommended by the statistician on the data safety monitoring committee; it decreases the potential for random error in the simple random selection approach.

Statistical Analysis

Simple counts and percentages were used to describe study outcomes within sibling groups. Chi-squared test and ANOVA were used to assess demographic differences between singletons and multiples.

We compared several methods for analysis of this clustered data. We first calculated odds ratios and confidence intervals for the primary outcome of the NO CLD study, survival without bronchopulmonary dysplasia, by logistic regression without adjustment for clustering. Then, we used two techniques using logistic regression that account for clustered data: generalized estimating equations and multiple outputation.[48,51, 62] Finally, we calculated a logistic regression based only on singletons and the first-enrolled infant from each multiple gestation. SAS version 9.1 was used for the analyses.

Results

Systematic Review

The search parameters yielded 79 papers, of which 41 met inclusion criteria. Only four (9%) of these (including the NO CLD trial) used statistics that accounted for the non-independence of multiples.[8,13,16,20] Collins et al. used logistic regression with robust variance estimates clustering on the mother.[16] Morley et al. presented logistic regression without clustering for the main analysis and stated that when the analysis was repeated with generalized estimating equations to account for multiple birth the "results were substantially unaffected."[20] Marret et al. used generalized estimating equations to account for the non-independence of the outcomes of twins; they also stratified randomization by multiple/singleton status and adjusted for this in the analysis, presumably to account for an unequal distribution of prognostic variables between multiples and singletons.[8] Eighteen papers did not specify whether multiples were included and four papers excluded all multiples or second-born infants from the study (figure 1). Of the six studies that described how enrolled multiples were randomized, five studies cluster-randomized the infants to the same intervention and one randomized the siblings separately. Among the postnatal intervention trials reporting the frequency of multiples in the study population, the prevalence ranged from 14–36%, although none of the studies specified how many of those subjects born from a multiple gestation had a sibling enrolled in the trial.

Case Study

Of the 582 infants enrolled in the NO CLD trial, 157 (23%) were twins or triplets. However, because some of these siblings were deceased or did not meet eligibility criteria, only 84 (14.4%) had a sibling enrolled in the trial. Among those with a sibling from the same gestation in the trial, there were 36 pairs of siblings (twins or two siblings from a higher-order multiple gestation), and 4 sets of three siblings (triplets or three of four quadruplets). In 26 of the sibling pairs (72%), both siblings had the same primary outcome (survival without bronchopulmonary dysplasia versus bronchopulmonary dysplasia or death). In all four sets of three siblings enrolled (100%), all the siblings shared the same outcome. In addition, statistically significant demographic differences were seen among singletons, multiples without a sibling enrolled in the trial, and multiples with a sibling enrolled in the trial (table 1). Multiples were more likely to be white and to be born to married parents and mothers of higher age (p<0.025).

The analysis of the primary outcome of the NO CLD trial showed sensitivity to the analytic approach to clustered data, with calculated odds ratios ranging from 1.36 to 1.52. In addition, statistical significance was inconsistent across analysis methods. Using simple logistic regression without adjusting for clustering resulted in a non-significant result while adjusting for clustering with either generalized estimating equations or multiple outputation resulted in statistical significance (p<0.05). Finally, when the odds ratio for the primary outcome was calculated using generalized estimating equations and multiple outputation, similar statistically significant estimates of the odds ratio and confidence intervals were obtained (table 2).

Discussion

We have demonstrated a case-study of an actual neonatal trial in which the conclusion of the study is sensitive to the analytic approach to multiples, and we have shown a low prevalence of papers accounting for such clustered data in the recent perinatal and neonatal literature. In the NO CLD trial, the outcomes of siblings from the same gestation were highly correlated as predicted based on shared genetic and environmental factors. Furthermore, likely as a result of a non-random societal distribution of *in vitro* fertilization,[63–65] infants from multiple gestations were different than singletons with respect to key demographic variables that may be predictors of important pulmonary and neurodevelopmental outcomes of prematurity.

Finally, whereas generalized estimating equations and multiple outputation yielded highly consistent estimates of a statistically significant effect, analysis by logistic regression without accounting for the clustering of outcomes did not reach significance.

The need for statistical approaches that account for clustering of data has been recognized for several decades.[45,47–49,52–54,56–58] Several papers in the 1990's showed that many cluster-randomized trials failed to account for clustered data in their analyses, leading to spurious results in more than 50% of such studies.[46,66,67] Our findings in the perinatal literature are that between 2003 and 2008, the majority of recent multi-center trials that measure outcomes in preterm infants have not accounted for multiples in their analyses. Furthermore, many do not address whether multiples were enrolled, and if so, how they were randomized. However, the use of generalized estimating equations in some of the trials reviewed may represent an early trend in the neonatal and perinatal literature.

Shared genetic and environmental influences certainly may cause the outcomes of multiples to be correlated. For instance, genetic effects account for approximately 80% of the observed variance in bronchopulmonary dysplasia susceptibility.[68] In addition, study designs that assign siblings to the same treatment may increase this correlation. Although independently randomizing each twin or systematically assigning them to different treatment arms could be one approach to decrease the correlation in postnatal trials, further work is needed to assess the statistical and ethical implications of this approach, in addition to the palatability to families. Certainly, in perinatal studies in which interventions are prenatal and the outcomes are measured in infants, this is not a feasible option. The exclusion of multiples from studies may also limit the generalizability of a trial, since multiples may be biologically and socio-demographically different than singletons. Therefore, this limitation should be seriously considered in the decision to exclude multiples from a study, as this approach could be more or less appropriate with different study aims. Furthermore, adjusting for multiple status in addition to clustering on pregnancy may be necessary in observational studies or in randomized trials when randomization has not successfully balanced the distribution of multiple status and associated prognostic covariates between the treatment groups. In situations where the outcomes of multiples and singletons are different, particularly if multiple-status is not a covariate in the model, multiple outputation may be more robust than GEE. Both multiple outputation and generalized estimating equations allow for adjusted analyses.

In mathematical simulations of a neonatal randomized clinical trial including twins, Shaffer et al. found only minimal differences between the results of logistic regression using generalized estimating equations and logistic regression without adjusting for correlated outcomes when the twins received the same therapy.[69] Gates et al., using one real and two simulated perinatal datasets, showed that confidence intervals are often wider with methods that account for the non-independence of multiples; using simulated datasets they showed that there is potential for inconsistency of point estimates of effect size using different methods. The NO CLD case study presents the example of an actual perinatal trial in which the non-independence of siblings' outcomes caused both the overall estimate of the odds ratio and confidence interval to be sensitive to the analytic approach, even with a lower percentage of twins than in the Gates analyses. Although the net differences were small, and some might argue the estimates were not meaningfully different, they would likely have been sufficient to alter many readers' interpretation of the trial results, since there was a difference in statistical significance. Analytic strategy that accounts for clustering will tend to generate the most conservative estimates of the confidence intervals and the most statistically valid point estimates of effect size.

Although logistic regression excluding all but the first-enrolled multiple was presented (table 2) to demonstrate the sensitivity of the calculated results to the handling of multiples, it is not a recommended approach because there are ethical concerns about excluding the data from

enrolled subjects and there is potential for systematic bias. For instance, in the NO CLD trial, infants enrolled earlier had a greater benefit from the study drug.[13] Therefore, systematically excluding data from infants enrolled later than their siblings could over-inflate the estimated odds ratio for benefit from therapy and does not make use of all of the gathered data. Potential bias exists both from excluding multiples from an analysis and also from eligibility in the trial itself. An "any occurrence" strategy, in which an outcome of the pregnancy is considered to have occurred if any of the multiples experiences the outcome, also fails to make use of all the information gathered in the study.[58] Randomly excluding data from analysis by randomly selecting one sibling from each pregnancy to contribute to the dataset similarly is prone to random error. Multiple outputation with a large number of repetitions is essentially an extension of this method that greatly decreases this possibility. However, multiple outputation is more computationally intensive and may be unfamiliar to readers. Generalized estimating equations may be more familiar to readers, but may also fail to converge (i.e. be statistically unstable) in situations with a low percentage of multiples. One reasonable strategy would be an *a priori* plan to use generalized estimating equations and, if they failed to converge, to use multiple outputation.[48] The Donner and Klar cluster trials method used by Gates et al. is also an option if adjustment for covariates is not necessary.[58,70] These approaches may also be expanded, for instance to allow for Bayesian modeling while accounting for the clustering. In addition, in some situations, multi-level clustering, such as by pregnancy and by center, may be appropriate.

While the case of the NO CLD trial demonstrates that the results of perinatal trials may be sensitive to the analytic approach to multiples, we have not determined the exact parameters under which clustered analyses are necessary; in general, the higher the percentage of multiples or the degree of intra-cluster correlation, the less valid the results of an analysis that assumes independence of outcomes will be. It is unclear what percentage of perinatal trials would have meaningfully different results with different statistical analyses. Therapies with small effect sizes, borderline significance, and few studies are most suspect, whereas therapies shown to have a large significant effect in many populations are of the least concern. Finally, although our search strategy for the systematic review of recent multi-center randomized clinical trials could have missed some existing trials, it is clear that the majority of recent multi-center trials have not accounted for the non-independence of multiples in their analyses. While we have focused on large randomized trials, similar analytical issues exist for other study designs, including cohort studies.[57]

In conclusion, while statistical approaches accounting for clustered data have become standard, [42] particularly in reports of trials in adults where the majority of outcomes are part of a cluster group, the results of this systematic review indicate that only a small minority of perinatal and neonatal trials measuring outcomes in premature infants account for the non-independence of multiples. The example of the NO CLD study demonstrates that the results of such studies, in which twins and triplets comprise a significant minority of the study population, can be sensitive to the analytic approach to clustered data. The degree of impact of clustering will depend on the number of multiple births in the sample and the degree to which there is correlation in the outcomes of multiples. This issue needs to be discussed *a priori* during trial design. Furthermore, trial reports should account for how multiples are enrolled and randomized. An analytic approach that accounts for the non-independence of siblings should be the default option, as clinical researches have a responsibility to both research subjects and future patients to present the most valid results possible.

Acknowledgments

References

1. August Fuhr N, Becker C, van Baalen A, Bauer K, Hopp H. Antibiotic therapy for preterm premature rupture of membranes - results of a multicenter study. J Perinat Med 2006;34(3):203–6. [PubMed: 16602839]

2. Dolitzky M, Inbal A, Segal Y, Weiss A, Brenner B, Carp H. A randomized study of thromboprophylaxis in women with unexplained consecutive recurrent miscarriages. Fertil Steril 2006;86(2):362–6. [PubMed: 16769056]

3. Hod M, Damm P, Kaaja R, Visser GH, Dunne F, Demidova I, et al. Fetal and perinatal outcomes in type 1 diabetes pregnancy: a randomized study comparing insulin aspart with human insulin in 322 subjects. Am J Obstet Gynecol 2008;198(2):186, e1–7. [PubMed: 17905178]

4. Als H, Gilkerson L, Duffy FH, McAnulty GB, Buehler DM, Vandenberg K, et al. A three-center, randomized, controlled trial of individualized developmental care for very low birth weight preterm infants: medical, neurodevelopmental, parenting, and caregiving effects. J Dev Behav Pediatr 2003;24 (6):399–408. [PubMed: 14671473]

5. Hake-Brooks SJ, Anderson GC. Kangaroo care and breastfeeding of mother-preterm infant dyads 0–18 months: a randomized, controlled trial. Neonatal Netw 2008;27(3):151–9. [PubMed: 18557262]

6. Sandri F, Ancora G, Lanzoni A, Tagliabue P, Colnaghi M, Ventura ML, et al. Prophylactic nasal continuous positive airways pressure in newborns of 28–31 weeks gestation: multicentre randomised controlled clinical trial. Arch Dis Child Fetal Neonatal Ed 2004;89(5):F394–8. [PubMed: 15321956]

7. Crowther CA, Hiller JE, Doyle LW, Haslam RR. Effect of magnesium sulfate given for neuroprotection before preterm birth: a randomized controlled trial. JAMA 2003;290(20):2669–76. [PubMed: 14645308]

8. Marret S, Marpeau L, Zupan-Simunek V, Eurin D, Leveque C, Hellot MF, et al. Magnesium sulphate given before very-preterm birth to protect infant brain: the randomised controlled PREMAG trial*. Bjog 2007;114(3):310–8. [PubMed: 17169012]

9. Peltoniemi OM, Kari MA, Tammela O, Lehtonen L, Marttila R, Halmesmaki E, et al. Randomized trial of a single repeat dose of prenatal betamethasone treatment in imminent preterm birth. Pediatrics 2007;119(2):290–8. [PubMed: 17272618]

10. Thornton JG, Hornbuckle J, Vail A, Spiegelhalter DJ, Levene M. Infant wellbeing at 2 years of age in the Growth Restriction Intervention Trial (GRIT): multicentred randomised controlled trial. Lancet 2004;364(9433):513–20. [PubMed: 15302194]

11. van der Ham DP, Nijhuis JG, Mol BW, van Beek JJ, Opmeer BC, Bijlenga D, et al. Induction of labour versus expectant management in women with preterm prelabour rupture of membranes between 34 and 37 weeks (the PPROMEXIL-trial). BMC Pregnancy Childbirth 2007;7:11. [PubMed: 17617892]

12. Askie LM, Henderson-Smart DJ, Irwig L, Simpson JM. Oxygen-saturation targets and outcomes in extremely preterm infants. N Engl J Med 2003;349(10):959–67. [PubMed: 12954744]

13. Ballard RA, Truog WE, Cnaan A, Martin RJ, Ballard PL, Merrill JD, et al. Inhaled nitric oxide in preterm infants undergoing mechanical ventilation. N Engl J Med 2006;355(4):343–53. [PubMed: 16870913]

14. Berseth CL, Van Aerde JE, Gross S, Stolz SI, Harris CL, Hansen JW. Growth, efficacy, and safety of feeding an iron-fortified human milk fortifier. Pediatrics 2004;114(6):e699–706. [PubMed: 15545616]

15. Carbone T, McEntire B, Kissin D, Kelly D, Steinschneider A, Violaris K, et al. Absence of an increase in cardiorespiratory events after diphtheria-tetanus-acellular pertussis immunization in preterm infants: a randomized, multicenter study. Pediatrics 2008;121(5):e1085–90. [PubMed: 18450851]

16. Collins CT, Ryan P, Crowther CA, McPhee AJ, Paterson S, Hiller JE. Effect of bottles, cups, and dummies on breast feeding in preterm infants: a randomised controlled trial. Bmj 2004;329(7459): 193–8. [PubMed: 15208209]

17. Dugas MA, Nguyen D, Frenette L, Lachance C, St-Onge O, Fougeres A, et al. Fluticasone inhalation in moderate cases of bronchopulmonary dysplasia. Pediatrics 2005;115(5):e566–72. [PubMed: 15833887]

18. Escobedo MB, Gunkel JH, Kennedy KA, Shattuck KE, Sanchez PJ, Seidner S, et al. Early surfactant for neonates with mild to moderate respiratory distress syndrome: a multicenter, randomized trial. J Pediatr 2004;144(6):804–8. [PubMed: 15192630]

19. Kelleher AS, Clark RH, Steinbach M, Chace DH, Spitzer AR. The influence of amino-acid supplementation, gestational age and time on thyroxine levels in premature neonates. J Perinatol 2008;28(4):270–4. [PubMed: 18288119]

20. Morley CJ, Davis PG, Doyle LW, Brion LP, Hascoet JM, Carlin JB. Nasal CPAP or intubation at birth for very preterm infants. N Engl J Med 2008;358(7):700–8. [PubMed: 18272893]

21. Moya FR, Gadzinowski J, Bancalari E, Salinas V, Kopelman B, Bancalari A, et al. A multicenter, randomized, masked, comparison trial of lucinactant, colfosceril palmitate, and beractant for the prevention of respiratory distress syndrome among very preterm infants. Pediatrics 2005;115(4): 1018–29. [PubMed: 15805380]

22. Ng SM, Turner MA, Gamble C, Didi M, Victor S, Malamateniou C, et al. TIPIT: a randomised controlled trial of thyroxine in preterm infants under 28 weeks gestation: magnetic resonance imaging and magnetic resonance angiography protocol. BMC Pediatr 2008;8:26. [PubMed: 18590560]

23. Steer P, Flenady V, Shearman A, Charles B, Gray PH, Henderson-Smart D, et al. High dose caffeine citrate for extubation of preterm infants: a randomised controlled trial. Arch Dis Child Fetal Neonatal Ed 2004;89(6):F499–503. [PubMed: 15499141]

24. Wang CL, Anderson C, Leone TA, Rich W, Govindaswami B, Finer NN. Resuscitation of preterm neonates by using room air or 100% oxygen. Pediatrics 2008;121(6):1083–9. [PubMed: 18519476]

25. Widness JA, Madan A, Grindeanu LA, Zimmerman MB, Wong DK, Stevenson DK. Reduction in red blood cell transfusions among preterm infants: results of a randomized trial with an in-line blood gas and chemistry monitor. Pediatrics 2005;115(5):1299–306. [PubMed: 15867038]

26. Anand KJ, Hall RW, Desai N, Shephard B, Bergqvist LL, Young TE, et al. Effects of morphine analgesia in ventilated preterm neonates: primary outcomes from the NEOPAIN randomised trial. Lancet 2004;363(9422):1673–82. [PubMed: 15158628]

27. Dani C, Bertini G, Pezzati M, Poggi C, Guerrini P, Martano C, et al. Prophylactic ibuprofen for the prevention of intraventricular hemorrhage among preterm infants: a multicenter, randomized study. Pediatrics 2005;115(6):1529–35. [PubMed: 15930213]

28. Field D, Elbourne D, Truesdale A, Grieve R, Hardy P, Fenton AC, et al. Neonatal Ventilation With Inhaled Nitric Oxide Versus Ventilatory Support Without Inhaled Nitric Oxide for Preterm Infants With Severe Respiratory Failure: the INNOVO multicentre randomised controlled trial (ISRCTN 17821339). Pediatrics 2005;115(4):926–36. [PubMed: 15805366]

29. Franz AR, Bauer K, Schalk A, Garland SM, Bowman ED, Rex K, et al. Measurement of interleukin 8 in combination with C-reactive protein reduced unnecessary antibiotic therapy in newborn infants: a multicenter, randomized, controlled trial. Pediatrics 2004;114(1):1–8. [PubMed: 15231900]

30. Jegatheesan P, Ianus V, Buchh B, Yoon G, Chorne N, Ewig A, et al. Increased indomethacin dosing for persistent patent ductus arteriosus in preterm infants: a multicenter, randomized, controlled trial. J Pediatr 2008;153(2):183–9. [PubMed: 18534218]

31. Lista G, Colnaghi M, Castoldi F, Condo V, Reali R, Compagnoni G, et al. Impact of targeted-volume ventilation on lung inflammatory response in preterm infants with respiratory distress syndrome (RDS). Pediatr Pulmonol 2004;37(6):510–4. [PubMed: 15114551]

32. Manzoni P, Stolfi I, Pugni L, Decembrino L, Magnani C, Vetrano G, et al. A multicenter, randomized trial of prophylactic fluconazole in preterm neonates. N Engl J Med 2007;356(24):2483–95. [PubMed: 17568029]

33. Moss RL, Dimmitt RA, Barnhart DC, Sylvester KG, Brown RL, Powell DM, et al. Laparotomy versus peritoneal drainage for necrotizing enterocolitis and perforation. N Engl J Med 2006;354(21):2225–34. [PubMed: 16723614]

34. Ng SC, Gomez JM, Rajadurai VS, Saw SM, Quak SH. Establishing enteral feeding in preterm infants with feeding intolerance: a randomized controlled study of low-dose erythromycin. J Pediatr Gastroenterol Nutr 2003;37(5):554–8. [PubMed: 14581796]

35. Osborn DA, Evans N, Kluckow M. Effect of early targeted indomethacin on the ductus arteriosus and blood flow to the upper body and brain in the preterm infant. Arch Dis Child Fetal Neonatal Ed 2003;88(6):F477–82. [PubMed: 14602694]

36. Ramanathan R, Rasmussen MR, Gerstmann DR, Finer N, Sekar K. A randomized, multicenter masked comparison trial of poractant alfa (Curosurf) versus beractant (Survanta) in the treatment of respiratory distress syndrome in preterm infants. Am J Perinatol 2004;21(3):109–19. [PubMed: 15085492]

37. Schulze A, Rieger-Fackeldey E, Gerhardt T, Claure N, Everett R, Bancalari E. Randomized crossover comparison of proportional assist ventilation and patient-triggered ventilation in extremely low birth weight infants with evolving chronic lung disease. Neonatology 2007;92(1):1–7. [PubMed: 17596730]

38. Simons SH, van Dijk M, van Lingen RA, Roofthooft D, Duivenvoorden HJ, Jongeneel N, et al. Routine morphine infusion in preterm newborns who received ventilatory support: a randomized controlled trial. JAMA 2003;290(18):2419–27. [PubMed: 14612478]

39. Sinha SK, Lacaze-Masmonteil T, Valls i Soler A, Wiswell TE, Gadzinowski J, Hajdu J, et al. A multicenter, randomized, controlled trial of lucinactant versus poractant alfa among very premature infants at high risk for respiratory distress syndrome. Pediatrics 2005;115(4):1030–8. [PubMed: 15805381]

40. Van Overmeire B, Allegaert K, Casaer A, Debauche C, Decaluwe W, Jespers A, et al. Prophylactic ibuprofen in premature infants: a multicentre, randomised, double-blind, placebo-controlled trial. Lancet 2004;364(9449):1945–9. [PubMed: 15567010]

41. Whitelaw A, Evans D, Carter M, Thoresen M, Wroblewska J, Mandera M, et al. Randomized clinical trial of prevention of hydrocephalus after intraventricular hemorrhage in preterm infants: brain-washing versus tapping fluid. Pediatrics 2007;119(5):e1071–8. [PubMed: 17403819]

42. Campbell MK, Elbourne DR, Altman DG. CONSORT statement: extension to cluster randomised trials. Bmj 2004;328(7441):702–8. [PubMed: 15031246]

43. Berlin JA, Kimmel SE, Ten Have TR, Sammel MD. An empirical comparison of several clustered data approaches under confounding due to cluster effects in the analysis of complications of coronary angioplasty. Biometrics 1999;55(2):470–6. [PubMed: 11318202]

44. Carlin JB, Gurrin LC, Sterne JA, Morley R, Dwyer T. Regression models for twin studies: a critical review. Int J Epidemiol 2005;34(5):1089–99. [PubMed: 16087687]

45. Cornfield J. Randomization by group: a formal analysis. Am J Epidemiol 1978;108(2):100–2. [PubMed: 707470]

46. Divine GW, Brown JT, Frazier LM. The unit of analysis error in studies about physicians' patient care behavior. J Gen Intern Med 1992;7(6):623–9. [PubMed: 1453246]

47. Donner A, Birkett N, Buck C. Randomization by cluster. Sample size requirements and analysis. Am J Epidemiol 1981;114(6):906–14. [PubMed: 7315838]

48. Follmann D, Proschan M, Leifer E. Multiple outputation: inference for complex clustered data by averaging analyses from independent data. Biometrics 2003;59(2):420–9. [PubMed: 12926727]

49. Goetgeluk S, Vansteelandt S. Conditional Generalized Estimating Equations for the Analysis of Clustered and Longitudinal Data. Biometrics. 2007

50. Hanley JA, Negassa A, Edwardes MD, Forrester JE. Statistical analysis of correlated data using generalized estimating equations: an orientation. Am J Epidemiol 2003;157(4):364–75. [PubMed: 12578807]

51. Liang K-YZS. Longitudinal data analysis using generalized linear models. Biometrika 1986;73:13–22.

52. Localio AR, Berlin JA, Ten Have TR, Kimmel SE. Adjustments for center in multicenter studies: an overview. Ann Intern Med 2001;135(2):112–23. [PubMed: 11453711]

53. Ten Have TR, Landis JR, Hartzel J. Population-averaged and cluster-specific models for clustered ordinal response data. Stat Med 1996;15(23):2573–88. [PubMed: 8961464]

54. Wears RL. Advanced statistics: statistical methods for analyzing cluster and cluster-randomized data. Acad Emerg Med 2002;9(4):330–41. [PubMed: 11927463]

55. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. Biometrics 1986;42(1):121–30. [PubMed: 3719049]

56. Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. Biometrics 1988;44(4):1049–60. [PubMed: 3233245]

57. Ananth CV, Platt RW, Savitz DA. Regression models for clustered binary responses: implications of ignoring the intracluster correlation in an analysis of perinatal mortality in twin gestations. Ann Epidemiol 2005;15(4):293–301. [PubMed: 15780777]

58. Gates S, Brocklehurst P. How should randomised trials including multiple pregnancies be analysed? Bjog 2004;111(3):213–9. [PubMed: 14961881]

59. Ballard RA. Inhaled nitric oxide in preterm infants--correction. N Engl J Med 2007;357(14):1444–5. [PubMed: 17914048]

60. Walsh MC, Wilson-Costello D, Zadell A, Newman N, Fanaroff A. Safety, reliability, and validity of a physiologic definition of bronchopulmonary dysplasia. J Perinatol 2003;23(6):451–6. [PubMed: 13679930]

61. Walsh MC, Yao Q, Gettner P, Hale E, Collins M, Hensman A, et al. Impact of a physiologic definition on bronchopulmonary dysplasia rates. Pediatrics 2004;114(5):1305–11. [PubMed: 15520112]

62. Hoffman EB, Sen PK, Weinberg CR. Within-cluster resampling. Biometrika 2001;88(4):1121–34.

63. Nachtigall RD. International disparities in access to infertility services. Fertil Steril 2006;85(4):871–5. [PubMed: 16580367]

64. Seifer DB, Frazier LM, Grainger DA. Disparity in assisted reproductive technologies outcomes in black women compared with white women. Fertil Steril. 2007

65. Wright VC, Chang J, Jeng G, Macaluso M. Assisted reproductive technology surveillance--United States, 2005. MMWR Surveill Summ 2008;57(5):1–23. [PubMed: 18566567]

66. Donner A, Brown KS, Brasher P. A methodological review of non-therapeutic intervention trials employing cluster randomization, 1979–1989. Int J Epidemiol 1990;19(4):795–800. [PubMed: 2084005]

67. Simpson JM, Klar N, Donnor A. Accounting for cluster randomization: a review of primary prevention trials, 1990 through 1993. Am J Public Health 1995;85(10):1378–83. [PubMed: 7573621]

68. Lavoie PM, Pham C, Jang KL. Heritability of bronchopulmonary dysplasia, defined according to the consensus statement of the national institutes of health. Pediatrics 2008;122(3):479–85. [PubMed: 18762515]

69. Choosing analytic methods for neonatal clinical trials that include twin births. Pediatric Academic Societies Meeting 2008; Honolulu, HI. E-PAS2008: 634457.16.

70. Donner, A.; Klar, N. Design and Analysis of Cluster Randomization Trials in Health Research. London: Arnold; 2000.
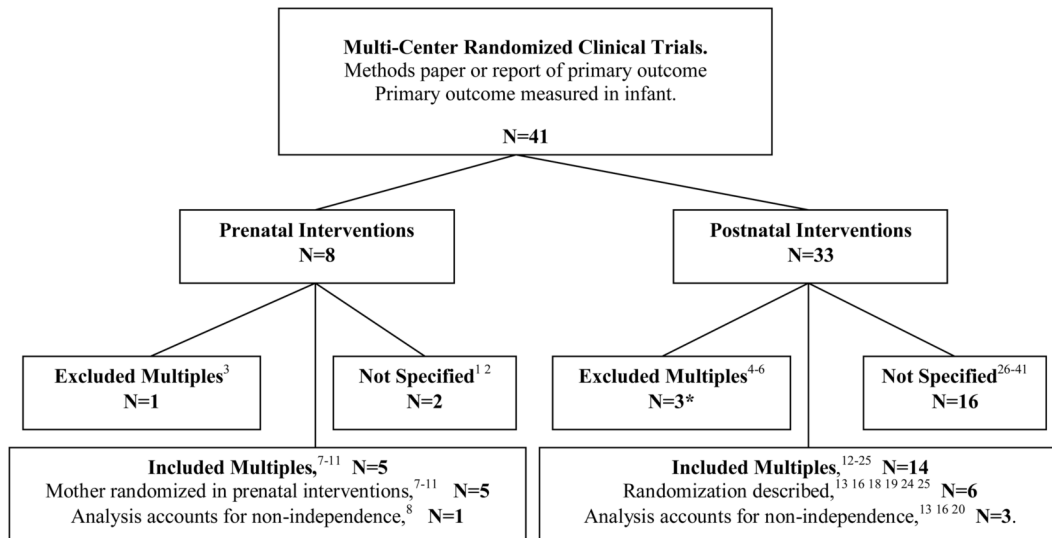
Figure 1. Results of systematic review
The majority of studies excluded multiples or did not specify whether they were enrolled. Only the minority of studies enrolling multiples used statistics that accounted for the non-independence of their outcomes. *Of the postnatal intervention trials that excluded multiples, two excluded all infants from a multiple gestation,[4,5] and one only included the first-born multiple.[6]

Table 1

Baseline demographic differences among infants enrolled in the NO CLD trial.

	Singleton (N=425)	Multiple: No sibling enrolled (n=73)	Multiple: Sibling enrolled (n=84)	p-value
Birth Weight (g) (mean ± SD)	762 ± 155	756 ± 186	769 ± 145	0.873
White (%)	49	67	71	0.001
Black (%)	31	23	19	
Hispanic (%)	15	6	7	
Parents Married (%)	50	69	69	0.025
Maternal Age (yr) (mean ± SD)	26.8 ± 6.6	29.2 ± 6.8	28.5 ± 5.6	0.004

Table 2

Odds ratios and 95% confidence intervals for survival without bronchopulmonary dysplasia with inhaled nitric oxide versus placebo calculated using different statistical approaches. Logistic regression excluding the second and third enrolled siblings includes all singletons and the first infant from any multiple gestation to be enrolled in the NO CLD trial.

Statistical Method	Odds Ratio (95% Confidence Interval)
Logistic Regression	1.36 (0.98–1.90)
Generalized Estimating Equations (GEE)	1.45 (1.03–2.04)
Multiple Outputation (MO)	1.42 (1.01–2.01)
Logistic Regression (excluding second and third enrolled siblings)	1.52 (1.08–2.15)