Randomised controlled trials for informing perinatal care

[1]Chris Gale, [2]William McGuire and [3]Edmund Juszczak

1. Section of Neonatal Medicine, Chelsea and Westminster campus, Imperial College London, United Kingdom

2. Centre for Reviews and Dissemination, University of York, York, UK

3. National Perinatal Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK

Corresponding author

Edmund Juszczak

National Perinatal Epidemiology Unit

Nuffield Department of Population Health

University of Oxford

Old Road Campus

Oxford

OX3 7LF

Tel: 01865 289700

Fax: 01865 289701

Email: ed.juszczak@npeu.ox.ac.uk

Keywords: Neonatal care, clinical trials, randomisation, pragmatic clinical trial

Main text word count: 2864 excluding abstract

ABSTRACT

Background

Randomised controlled trials provide the best evidence for the effects of interventions and are a key tool in efforts to improve the care and outcomes for newborn infants.

Methods

We discuss the role of randomisation for minimising selection bias using the play of chance in trials, and describe examples of seminal trials that have shaped the development of modern perinatal care. We consider the challenges inherent in developing and delivering large, simple and pragmatic trials, and the need for the development and adoption of core outcome sets to ensure that trials provide high-quality evidence of sufficient validity and applicability to guide policy and practice.

Results

Since the earliest days of modern neonatology, the randomised controlled trial has been recognised as the best method for assessing treatments and practices. While many strategies that reduce mortality and morbidity have been introduced following randomised trials, there are important examples of ineffective or potentially harmful practices that have been adopted in the absence of trial-based evidence. Typically, randomised controlled trials in perinatal care need to recruit several thousand participants to be able to detect modest but potentially important effects of new interventions on the most important but rare outcomes. Given the concerns about the financial burden and regulatory complexity of standard trial designs, innovative “efficient” trial designs are being evaluated to streamline processes while safeguarding participants.

Conclusions

Well conducted randomised controlled trials provide the most robust evaluation of interventions aimed at improving outcomes for newborn infants and their families. Increasingly, these trials will need to be large, multicentre (often international), use a simple and pragmatic protocol, incorporating meticulous follow-up procedures and assessment of long-term outcomes.

Well-conducted randomised controlled trials (RCTs) are essential for assessing interventions to improve outcomes for preterm or sick newborn infants. High-quality RCTs are the only way that causality, rather than association, can be demonstrated reliably. Seminal RCTs have heralded the widespread adoption of some of the most impactful treatments in neonatology including exogenous surfactant replacement for very preterm infants with respiratory distress syndrome (1), therapeutic hypothermia for term infants with neonatal encephalopathy (2), and simple delivery room interventions to prevent hypothermia in low birth weight infants (3).

Origins of trial-based evidence

One of the earliest controlled trials using random allocation in perinatal care was reported by William Silverman, a pioneering advocate of evidence-based practice, and colleagues in New York in 1952 (4). The investigators assessed the effect of corticitropin (adrenocorticotropic hormone) on disease progression in low birth weight infants with (what is now called) retinopathy of prematurity. Treatment versus expectant management was allocated by drawing marbles from an opaque jar making this (probably) the first truly randomised trial in neonatology. Analysis did not show any treatment benefits but demonstrated higher mortality in infants allocated to receive corticitropin. This seminal study illustrates how the young discipline of neonatology aimed to assess new interventions using a "fair test"; however, many other examples exist where interventions were introduced without supporting trial-based evidence, sometimes with adverse consequences that were not revealed for many years (5). A salutary example is the association of supplemental oxygen with retinopathy of prematurity in preterm infants (6). Oxygen therapy became widespread in newborn care from the 1940s and this was followed by an epidemic of "infantile blindness". A prospective multi-centre randomised trial was undertaken in 1953 to compare the effect of "routine oxygen" with "curtailed oxygen" on outcomes including mortality, retinopathy of prematurity and neurological status (although assessment of this outcome was never undertaken). On initial analysis, this trial showed lower levels of retinopathy of prematurity in the curtailed oxygen group without an effect on mortality prompting widespread adoption of strict upper limits on oxygen exposure for preterm infants. Unrecognised at the time, however, the trial had important methodological flaws. Infants who died within 48 hours after birth were not included in the analyses, concealing the fact that early mortality was much higher in the curtailed oxygen group. Subsequent analysis estimated that for each infant in the curtailed oxygen who avoided visual impairment, several other infants may have died (7). The clinical uncertainty that followed took more than 60 years to be addressed adequately through large international collaborative randomised controlled trials (8). This saga highlights the complexity and difficulties inherent in designing such practice-defining trials, as well as serving as reminder of their irreplaceable role in identifying optimal neonatal care.

Bias in RCTs

"The RCT is a very beautiful technique, of wide applicability, but as with everything else there are snags"- Archie Cochrane 1972

Randomisation minimises the effects of bias using the play of chance; it is the only means of controlling for unknown confounding factors in addition to those that are known and measured. However, potential sources of bias still exist within RCTs and these are important to consider when assessing the validity and applicability of trial findings (Table 1) (9).

Table 1. Cochrane Risk of Bias tool for RCTs (9)

• *Selection bias* due to inadequate sequence generation or inadequate allocation concealment
• *Performance bias* due to inadequate masking of participants/clinicians/investigators
• *Detection bias* due to inadequate masking of outcome assessors
• *Attrition bias* due to incomplete outcome data
• *Reporting bias* due to only selected outcomes being reported
• *Other forms of bias*, such as imbalance between groups in baseline characteristics

Selection bias

Although the size and direction of the effect of the different sources of bias on outcomes varies between trials, empirical evidence indicates that inadequate allocation concealment is a key threat to trial validity (10). Allocation concealment, traditionally achieved by storing the randomisation sequence in sealed, opaque envelopes, is now commonly achieved using remote web- or telephone-based services. Concealing allocation protects against selection bias by preventing investigators from being able to predict each participant's treatment group. Predictability subverts randomisation and can introduce systematic differences in the prognosis of the groups that are being compared. Appropriately concealed randomisation still allows differences between groups at baseline to occur by chance but this is minimised by enrolling a sufficiently large number of participants and the use of stratification. Reassuringly, a recent analysis of the methodological quality and risk of bias in 848

neonatal trials included in Cochrane Neonatal systematic reviews indicates that the proportion of trials that were rated as being at high risk of selection bias has decreased progressively over the past six decades (11).

Sample size matters

The introduction of antenatal steroids and exogenous surfactants, some of the most important advances in care for very preterm infants over the past 50 years, were informed by data from RCTs and meta-analyses (12) demonstrating that these interventions lead to a more than 40% reduction in mortality and other adverse outcomes. It is highly unlikely, however, that new and emerging interventions for newborn infants will have the same major beneficial effects since progress is typically incremental. Consequently, trials must seek to detect much smaller effect sizes, for example, a 10% relative reduction in the risk of mortality and other adverse outcomes. A major issue facing investigators undertaking such RCTs is recruiting sufficient numbers of infants to meet the statistical "sample size" requirements to detect modest but clinically relevant effects on uncommon outcomes (Table 2).

Table 2. Examples of sample size estimates (90% power, 2-sided significance level of 5%, with no attrition)

Target effect size (relative risk reduction)	**Change in event rate**	**Sample size per group**
20%	10% to 8%	> 4,300
15%	10% to 8.5%	> 7,800
10%	10% to 9%	> 18,000

An otherwise methodologically robust RCT is unlikely to be informative if it enrols too few participants to allow statistical analyses to generate estimates of effect that are precise enough to inform policy or practice decisions (13). Despite this requirement, historically, most published RCTs of neonatal care interventions have enrolled fewer than about 100 participants (14). Although small RCTs may not independently provide data sufficient to guide practice, this does not invalidate the trial findings or indicate that the report should not be published. In the era of research synthesis, trials do not need to be stand alone as independent works, but they do need to be of high quality so that the data can feed usefully into systematic reviews and meta-analyses. However, small RCTs do tend to have more design limitations than large trials, and are more subject to publication bias, that is, their results are more likely to be disseminated if there is evidence of a treatment effect. Concerns exist, therefore, that inclusion of their data in meta-analyses might yield biased over-estimates of treatment effects (15).

Recent instances where data from large trials have overturned previous estimates of effect generated from smaller trials illustrate this concern. For example, a series of small RCTs that assessed intravenous immunoglobulin for treating preterm infants with serious infection were identified, appraised and synthesised in a Cochrane systematic review (16). Individually, these trials were too small to detect meaningful treatment effects reliably, but meta-analysis of their combined data showed a seemingly beneficial effect of intravenous immunoglobulin on outcomes including mortality. This evidence, however, was not considered robust enough to inform practice because of the risks of bias inherent in small trials. Consequently, investigators undertook a large simple pragmatic RCT to resolve the uncertainty (17). The "INIS" international multi-centre trial, one of the largest individually randomised neonatal trials undertaken, found no evidence of benefit for intravenous immunoglobulin and overturned the previous evidence-base so that the updated Cochrane review including INIS data was able to state conclusively that intravenous immunoglobulin is not effective in this context (Figure 1).

Challenges in undertaking large, pragmatic RCTs

While RCTs in which adults with common conditions such as cardiovascular disease or diabetes participate may be feasibly conducted within a few clinical centres, large RCTs in neonatology must involve many centres, sometimes multi-nationally. Individually, neonatal centres provide care for relatively few infants in the highest risk groups, for example extremely preterm, or for infants with relatively rare conditions, for example neonatal encephalopathy. To be feasible, trials that aim to recruit thousands of participants have to involve tens if not hundreds of neonatal units. The INIS trial, for example, enrolled 3,493 infants in 113 hospitals in nine countries (17). Without this scale of collaboration, RCTs in a few sites would need to continue recruitment for many years to enrol the required number of participants, which is impractical.

The role of clinical trial units

Increasingly, large multi-centre RCTs are conducted within collaborative research networks of perinatal units. Co-ordination of such complex and costly endeavours is best managed by staff in dedicated *clinical trials units* with the expertise to provide the statistical, epidemiological,

logistical and methodological advice needed to develop and deliver such trials efficiently and successfully (18). Clinical trials units can maintain the secure information technology frameworks and robust quality assurance systems required to ensure participant data protection and confidentiality. This managed multi-disciplinary approach can also facilitate early and meaningful involvement of infant- and family-advocates (19). Such co-ordinated engagement allows parent groups and researchers to prioritise their most important questions and outcomes and to ensure that competing trials do not occur simultaneously.

Pragmatic trials

Large multi-centre RCTs are more likely than single-centre trials to change clinical practice. A key feature of these trials is that they can be considered pragmatic, that is, trial processes (enrolment, randomisation and allocation, outcome assessment and reporting) are performed by clinicians (nurses, midwives, paediatricians and neonatologists) in a representative spectrum of centres, and interventions are tested and evaluated in the context of existing clinical care pathways using infant- and family-relevant, routinely measured outcomes (20). This pragmatism enables trials to recruit the numbers of infants needed to evaluate clinically important effect sizes. Clinicians and service-users engage more fully when trials are likely to produce results applicable to their circumstances and generalizable across different healthcare settings and types of neonatal units.

Assessing meaningful outcomes

To be valid and applicable, and to inform policy and practice, RCTs need to measure meaningful (sometimes modest) effects on the outcomes of greatest importance to infants, their families, carers and health services. Historically, the most important outcome for sick and preterm newborn infants has been mortality. As care practices have advanced and survival chances for sick and preterm infants have improved, assessing the effect of interventions on other morbidities, particularly adverse neurodevelopmental outcomes, has become increasingly important. This is particularly relevant in neonatology as perinatal interventions may improve short term outcomes, including survival, but at a cost of adverse longer-term outcomes. For example, giving very preterm infants systemic dexamethasone in the first few days after birth facilitates earlier weaning from mechanical ventilation and reduces supplemental oxygen requirements, but on long term follow-up however infants who received dexamethasone had a higher rate of adverse neurological effects, including cerebral palsy (21).

Measuring longer term growth, neurological and developmental outcomes is logistically challenging and expensive, far more so than measuring morbidity or mortality during initial hospitalisation. Some neurodevelopmental assessment tools are validated and applicable for use beyond about 18 months post-term, but assessment of milder problems and of cognitive function and behavioural and educational outcomes can only reliably be undertaken after children are about 5 years old (22). It is important that neurodevelopmental assessment is achieved for as many of the trial participants as possible. Incomplete follow up of the trial cohort may be an important source of bias because participants who are assessed at long term follow up tend to have fewer and less severe adverse neurodevelopmental outcomes than those in whom neurodevelopmental assessments are not undertaken (for example, because the investigators were not able to see the child in an assessment centre) (23).

Although disability-free survival is rightly regarded as an important trial outcome for informing practice decisions, other outcomes are sometimes useful in specific contexts. For example, the “TIPP trial” showed that prophylactic indomethacin does not affect survival without neurosensory impairment at 18 months in extremely low birth weight infants (24). However, the trial did show that prophylactic indomethacin reduced the likelihood that infants received surgery to close the patent ductus arteriosus. The decision on whether to use prophylactic indomethacin depends on the values attached by parents and carers to these benefits. For example, a reduced need for surgery may be of particular importance in centres where infants need to be moved to specialist cardiology and thoracic surgery centre.

Composite outcomes

One consequence of the desire to ensure that all important outcomes are accounted for has been the relatively widespread use of composite outcomes in perinatal trials, with an additional perceived advantage in that combined outcomes necessitate lower sample sizes, due to a greater number of expected events. Composite outcomes include two or more (usually rare) outcomes that are considered to be important. For example, RCTs of deferred cord clamping for very preterm infants specified a primary outcome consisting of five components: death; severe brain injury shown on ultrasonography; severe retinopathy of prematurity; necrotising enterocolitis; and late-onset infection (25). Some outcomes are included within the composite because they are competing: for example, infants who die during the early neonatal period will not develop retinopathy, and the composite accounts for this. Furthermore, use of a composite primary outcome increases the statistical “power” of an RCT, that is, its ability to detect a real effect, if it exists. Although statistically efficient, however, such combined endpoints can be difficult to

interpret, particularly where the different components might not be considered equally important. This problem is compounded in trials where treatment effects are observed working in opposite directions for the components of the composite as seen in the SUPPORT trial of oxygen saturation targeting which showed that the risk of severe retinopathy of prematurity was significantly reduced, while that of death before discharge was increased, in preterm infants allocated to the lower oxygen saturation target range (26).

Core outcome sets

Concerns exist across all fields of medicine that the misuse of surrogate outcomes that do not directly inform policy or practice leads to inefficient and wasteful research (27). Trials reporting outcomes that are not relevant to infants, parents or health services do not inform clinical practice. Furthermore, use of multiple different outcomes measures leads to heterogeneity across trials so limiting the validity of meta-analyses of their combined data. One solution to this problem is the development and uptake of "core outcome sets": a *minimum* set of outcomes developed in partnership with health professionals, patients and parents, to be measured in all trials in a standardised manner. The identification and use of such core outcomes in neonatology (29) should reduce research waste by facilitating evidence synthesis and address the increasingly recognised concern that parents and ex-neonatal patients identify and rank different outcomes as being important when compared to clinicians and researchers (27, 28).

More efficient trial designs

In many countries with integrated health services, developments in clinical practice and research are increasingly being driven by the ubiquity of electronic medical record systems and the application of artificial intelligence to "big" and "real world" data. Embracing these developments provides opportunities to ensure that RCTs maintain their central role guiding and improving perinatal care practices, and are able to do so more rapidly and cost-effectively. In particular, use of electronic data capture systems (such as electronic patient records and registries) provides an opportunity to develop and conduct very large RCTs, so called "mega-trials", by making processes simpler, cheaper and more efficient (30). For example, in the UK, the "WHEAT trial" of withholding enteral feeds during packed red cell transfusion to prevent necrotising enterocolitis in preterm neonates (https://www.npeu.ox.ac.uk/wheat) is pioneering the integration of all trial processes: identification of eligible participants, randomisation, recording consent and trial data capture all occur within existing electronic patient record systems and research databases. This approach makes incorporating trial processes into day-to-day practice across hundreds of neonatal units potentially possible and provides the opportunity for involvement in neonatal clinical trials to be the norm, rather than the exception. Such trials require efficient recruitment and data collection, and simplified approaches to consent; it is reassuring that research ethics committees consider this approach consistent with good participant's care and trial governance (31).

Conclusion

Over the past 60 years, neonatal health professionals and researchers in collaboration with parents and families have realised enormous improvements in care and outcomes for sick and preterm infants, and these advances have often been heralded by high quality RCTs. Notwithstanding these successes, many day-to-day aspects of neonatal care still lack high quality evidence to inform practice. Since RCTs are the best way to assess causality, they will continue to be the bedrock upon which high quality evidence based care is built, and the best option for addressing existing practice uncertainties and those that arise in future. Technological and methodological advances offer great potential to make high quality randomised trials simpler, easier and more inclusive – such trials will be essential if we are to continue improving the care and outcomes of babies that need neonatal care.

Acknowledgements

We thank Sir Iain Chalmers and the fascinating records made publicly available at the James Lind Library (www.jameslindlibrary.org).

References

1. Curstedt T, Halliday HL, Speer CP. A unique story in neonatal research: the development of a porcine surfactant. Neonatology 2015;107:321-9.

2. Azzopardi DV, Strohm B, Edwards AD, Dyet L, Halliday HL, Juszczak E, et al. Moderate hypothermia to treat perinatal asphyxial encephalopathy. N Engl J Med 2009;361:1349-58.

3. McCall EM, Alderdice F, Halliday HL, Vohra S, Johnston L. Interventions to prevent hypothermia at birth in preterm and/or low birth weight infants. Cochrane Database Syst Rev 2018;2:CD004210.

4. Reese AB, Blodi FC, Locke JC, Silverman WA, Day RL. Results of use of corticotropin (ACTH) in treatment of retrolental fibroplasia. AMA Arch Ophthalmol 1952;47:551-5.

5. Halliday HL. Useless perinatal therapies. Neonatology 2010;97:358-65.

6. Silverman WA. A cautionary tale about supplemental oxygen: the albatross of neonatal medicine. Pediatrics 2004;113(2):394-6.

7. Bolton DP, Cross KW. Further observations on cost of preventing retrolental fibroplasia. Lancet 1974;1:445-8.

8. Askie LM, Darlow BA, Finer N, Schmidt B, Stenson B, Tarnow-Mordi W, et al. Association Between Oxygen Saturation Targeting and Death or Disability in Extremely Preterm Infants in the Neonatal Oxygenation Prospective Meta-analysis Collaboration. JAMA 2018;319:2190-2201.

9. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al; Cochrane Bias Methods Group; Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928.

10. Odgaard-Jensen J, Vist GE, Timmer A, Kunz R, Akl EA, Schünemann H, et al. Randomisation to protect against selection bias in healthcare trials. Cochrane Database Syst Rev 2011;4:MR000012.

11. Lai N, Ong J, Chaiyakunapruk N, Chen K, Ovelman C, Soll R. Are neonatal trials better conducted and reported over the last six decades? An analysis on their risk-of-bias status in Cochrane reviews. Neonatology 2019 [in press]

12. Sweet DG, Carnielli V, Greisen G, Hallman M, Ozek E, Plavka R, et al. European consensus guidelines on the management of respiratory distress syndrome - 2016 update. Neonatology 2017;111:107-125.

13. Foglia EE, DeMauro SB, Dysart K, Kirpalani H. When has enough evidence accumulated to change neonatal practice? Semin Fetal Neonatal Med 2015;20:424-30.

14. DeMauro SB, Giaccone A, Kirpalani H, Schmidt B. Quality of reporting of neonatal and infant trials in high-impact journals. Pediatrics 2011;128:e639-44.

15. Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K. Publication bias in clinical trials due to statistical significance or direction of trial results. Cochrane Database Syst Rev 2009:MR000006.

16. Ohlsson A, Lacy JB. Intravenous immunoglobulin for suspected or proven infection in neonates. Cochrane Database Syst Rev 2015;(3):CD001239.

17. INIS Collaborative Group, Brocklehurst P, Farrell B, King A, Juszczak E, Darlow B, et al. Treatment of neonatal sepsis with intravenous immune globulin. N Engl J Med 2011;365:1201-11.

18. Gale C, Juszczak E. A paediatrician's guide to clinical trials units. Arch Dis Child Educ Pract Ed 2016;101:265-7.

19. Janvier A, Bourque CJ, Dahan S, Robson K, Barrington KJ; on behalf of the Partenariat Famille (PAF) team. Integrating parents in neonatal and pediatric research. Neonatology 2019;115:283-291.

20. Ford I, Norrie J. Pragmatic Trials. N Engl J Med 2016;375:454-63.

21. Halliday HL. Update on postnatal steroids. Neonatology 2017;111:415-422.

22. Marlow N. Measuring neurodevelopmental outcome in neonatal trials: a continuing and increasing challenge. Arch Dis Child Fetal Neonatal Ed 2013;98:F554-8.

23. Tin W, Fritz S, Wariyar U, et al. Outcome of very preterm birth: children reviewed with ease at 2 years differ from those followed up with difficulty. Arch Dis Child Fetal Neonatal Ed 1998;79:F83–7.

24. Schmidt B, Davis P, Moddemann D, Ohlsson A, Roberts RS, Saigal S, et al, Trial of indomethacin prophylaxis in preterms I. Long-term effects of indomethacin prophylaxis in extremely-low-birth-weight infants. N Engl J Med 2001; 344: 1966–72.

25. Tarnow-Mordi W, Morris J, Kirby A, Robledo K, Askie L, Brown R, et al. Delayed versus immediate cord clamping in preterm infants. N Engl J Med 2017;377:2445-55.

26. SUPPORT Study Group of the Eunice Kennedy Shriver NICHD Neonatal Research Network, Carlo WA, Finer NN, Walsh MC, Rich W, Gantz MG, et al. Target ranges of oxygen saturation in extremely preterm infants. N Engl J Med. 2010;362:1959-69.

27. Gale C, Modi N. Towards greater efficiency in neonatal clinical research. Lancet Child Adolesc Health 2017;1:169-70.

28. Webbe J, Brunton G, Ali S, Longford N, Modi N, Gale C, et al. Parent, patient and clinician perceptions of outcomes during and following neonatal care: a systematic review of qualitative research. BMJ Paediatr Open 2018;2:e000343.
29. Webbe J, Brunton G, Ali S, Duffy JMN, Modi N, Gale C. Developing, implementing and disseminating a core outcome set for neonatal medicine. BMJ Paediatrics Open 2017;1(e000048).
30. Gale C, Modi N. Towards greater efficiency in neonatal clinical research. Lancet Child Adolesc Health 2017;1:169-70.
31. Gale C, Hyde MJ, Modi N; WHEAT trial development group. Research ethics committee decision-making in relation to an efficient neonatal trial. Arch Dis Child Fetal Neonatal Ed 2017;102:F291-F298.

Figure 1. Meta-analysis of RCTs of intravenous immunoglobulin for suspected or proven infection in neonates- INIS overturns previous evidence-base (16)