
Title: Continued uncertainty regarding treatment of Patent Ductus Arteriosus in
premature infants and the role of clinical trials

First Author Name, Qualifications: Edmund Juszcak¹, BSc(Hon.), MSc, CStat, CSci
Job Title: Director, National Perinatal Epidemiology Unit (NPEU) Clinical Trials Unit

Author Name, Qualifications: Samir Gupta^{2,3} MBBS, DM, MRCP, MD, FRCPCH, FRCPI
Job Title: Professor of Neonatology, Durham University Honorary Research Fellow, University of Oxford Director, Research & Development Consultant Neonatologist

Author's Institutional Affiliations:

1. NPEU Clinical Trials Unit, Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom (ed.juszcak@npeu.ox.ac.uk)
2. Department of Paediatrics & Neonatology, University Hospital of North Tees, Stockton-on-Tees, United Kingdom, (samir.gupta@nth.nhs.uk)
3. Durham University, Stockton-on-Tees, United Kingdom.

Corresponding Author:

Edmund Juszcak
NPEU Clinical Trials Unit
National Perinatal Epidemiology Unit
Nuffield Department of Population Health
University of Oxford
Old Road Campus, Headington
OXFORD, OX3 7LF
United Kingdom
Email: ed.juszcak@npeu.ox.ac.uk
Tel: +44 - (0)1865 289700

Postal address for Samir Gupta:

University Hospital of North Tees
Stockton-on-Tees,
TS19 8PE
United Kingdom

Email: samir.gupta@nth.nhs.uk

Disclosure statement: No conflict of interest to disclose.

Category of study: Review

Word count of abstract: 136

Word count of manuscript: 5,027 excluding references & abstract (555 words in the references)

Number of Figures: 1

Number of Tables: None

Key words: Preterm, Patent Ductus Arteriosus (PDA), Intraventricular Haemorrhage (IVH), Ductal Closure, Randomised, Care Pathway, Uncertainty, Trials, Evidence.

Abstract

Despite several decades of research into treatments for patent ductus arteriosus (PDA), there is continued uncertainty regarding if, when and how best to treat a PDA and the long term consequences. There are almost 5,000 babies enrolled into clinical trials but the questions remain largely unanswered. Many of the trials performed over the period were well designed and addressed important clinical questions, but the results are not necessarily directly applicable to the clinical management dilemmas of today since perinatal care has improved over time per se, the patient population is typically more premature and there have been technological advances in diagnosis. In the article we examine some of the approaches taken, how trial designs evolved over time, especially in terms of the patient population and outcomes evaluated, and offer points to consider when planning future research.

A lesson from history

A good place to start our critique, by analogy, is the first ever reported multi-arm trial, conducted by James Lind, a Scottish physician. In the 18th Century sailors were dying from scurvy; many different treatments were being used to try to treat the disease. Lord Anson, the Admiral of the fleet, reported in 1748 that 380 out of 510 crew on one of his ships died of the disease. According to Lind, scurvy caused more deaths in the British fleets than French and Spanish arms.¹ Lind quickly realised that “No physician conversant with this disease at sea had undertaken to throw light upon the subject”. So in May 1747, he conducted on 12 patients with the scurvy, on board the Salisbury at Sea. Their cases were clearly described - they all had putrid gums, the spots and lassitude, with weakness of their knees (Patients). They were housed together in one place (setting), in a proper apartment for the sick in the fore-hold and had a common diet. Six different interventions (I) were given to 2 patients respectively. These ranged from oranges and lemons to sea water and other concoctions of varying acidity. The results appeared conclusive - one of the men who had taken oranges and lemons was fit for duty after six days, the other also making progress.

What were the strengths and weaknesses of this trial?

The strengths include the description of and homogeneity of the patient population, the control of diet and movements and the use of an objective substantive outcome (good for the captain at least).

The weaknesses include the tiny sample size, the undocumented process of randomisation, the generalisability of the findings and potential selection bias; two patients with the most severe symptoms were reported to have received sea-water.

What can we learn from this trial?

Good design and conduct is essential; clarity of definitions, the standardisation of approach and the use of objective outcomes all help to 'hear a signal above the noise', plus sometimes you get lucky.

A review of the 'PDA trials' literature

- i. Using an example of a trial published in 1983 studying the effects of indomethacin in premature infants with PDA ²

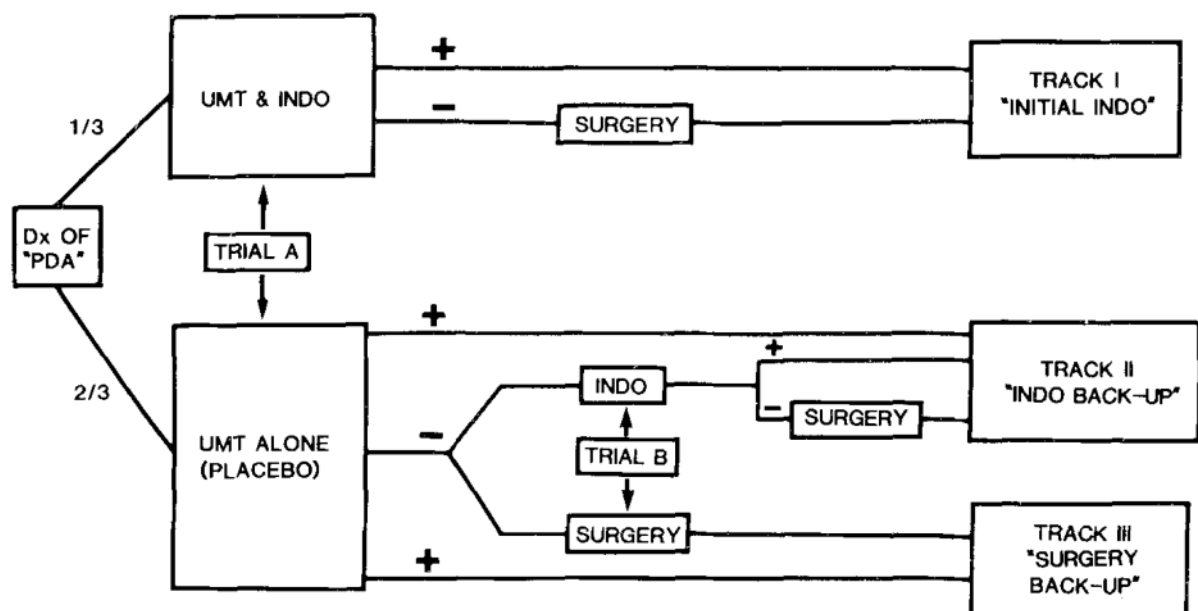
The design ethos was particularly impressive in terms of deliverability – the trial was designed to “simulate the therapeutic options available to clinicians caring for a preterm infant who develops a serious, potentially life-threatening PDA”; a care-pathway trial in modern parlance. The participants (**P**) were 421 new-born infants with birth weight <1,750g who had developed a hemodynamically significant PDA. One of three regimens (**I**) could be pursued (Figure 1):

Care-pathway 1: attempt to reduce the effects of volume overload on the cardiovascular system using an intensive course of 'Usual Medical Therapy' (UMT entailed e.g., fluid restriction, diuretics, and perhaps digoxin) and at the same time administer indomethacin (UMT + INDO).

Care-pathway 2: Perform UMT as in described above, but if this approach is not effective within a short period of time (initial re-evaluation at 36 to 48 hrs) in ameliorating the clinical impact of the PDA, administer indomethacin.

Care-pathway 3: Perform UMT as described above, but if this approach is not effective within a short period of time in ameliorating the clinical impact of the PDA, refer the infant for surgical ligation of the ductus arteriosus.

Figure 1. Schema of study design, summarising randomisation allocations on diagnosis of haemodynamically significant PDA in trial A and subsequent randomisation in trial B of infants initially receiving placebo who required backup therapy.



The primary outcome (O) was ductal closure rate – now considered a surrogate outcome but clearly relevant since the drug is being given with the aim of closing the duct. Secondary outcomes included hospital mortality, incidence of adverse conditions during hospitalisation, length of hospital stay and duration of respiratory support.

This was an elegant but complex design using a two-stage randomisation which addressed two important research questions.

1. Initial treatment (Trial A) allowed a comparison of 'early treatment' (up to 48 hours) in indomethacin vs placebo in a targeted high risk group.

2. Trial B allowed a comparison of indomethacin vs surgery in those infants for whom the PDA had not closed by itself within 36-48 hours, a 'later treatment' clinical question.

The main comparative analysis was performed on an intention-to-treat basis (ITT) i.e. infants were analysed in their randomised groups irrespective of treatment received. After 48 hours of treatment the PDA closure rate was 79% in the indomethacin group vs 28% in the placebo group. This control group event rate rose to 35% without the need for backup therapy, and this comparison was reported as the primary outcome (RR 2.3 $P < 0.001$). Another finding was that closure rate was related to gestational age.

What were the strengths and weaknesses of this trial?

Strengths include the elegance of the design, the clarity of reporting, the trial addressed important (multiple) clinical questions including safety, was embedded in routine clinical practice and used a multidisciplinary team. It was perhaps ahead of its time, and such a design would be most welcome today. Weaknesses include the complexity (could also be seen as a strength), a lack of a sample size/power calculation, the analysis and interpretation (hypothesis tests did not always compare like with like and the equivalence conclusion regarding the timing of administration of indomethacin is questionable) and multiple subgroup analysis was given too much prominence.

What can we learn from this trial?

The bottom line is that embedding a trial within all present treatment options in care pathways will maximise the chances of success by ensuring maximum clinician buy-in. Addressing more than one research question makes the trial not only important but cost-effective.

- ii. Using an example of a trial published in 1994 studying low-dose indomethacin and prevention of intraventricular haemorrhage.³

The aim was to test the hypothesis that low-dose indomethacin would lower the incidence and severity of intraventricular haemorrhage (IVH). The participants (**P**) were 431 neonates of birth

weight 600–1,250g with no evidence for IVH at 6 to 11 hours of age. The intervention (**I**) was low-dose indomethacin (0.1mg/kg IV at 6 to 12 postnatal hours and every 24 hrs for two more doses) or (comparator **C**) an equal volume of saline placebo by slow intravenous infusion during a 5- to 10-minute period.

The primary outcome (**O**) was not explicitly stated but the sample size was based on 'IVH rate', clearly a more substantive outcome than PDA closure with potential life-long consequences.

Secondary short-term outcomes included PDA closure, haemorrhage, bronchopulmonary dysplasia (BPD), mortality and adverse reactions.

What were the strengths and weaknesses of this trial?

Strengths include a sample size/power re-calculation during the trial due to a lower than expected control group event rate; furthermore they very nearly achieved their re-estimated sample size target (19 short but the original sample size was 250–300). The trial was multi-centre, used a robust randomisation method/system, the clarity of reporting was evident in the definitions and potential 'reproducibility' of the trial. There was also very low attrition and good safety reporting.

The main weaknesses were regarding some of the reporting and analytical methods. There was no clear specification of the primary outcome and possible selective reporting e.g. in the abstract both IVH and mortality were reported in two different ways respectively. One could question the analytical methods used e.g. the analysis of nominated primary outcome used a test for trend.

Typically, if a test for trend is used, the increments are expected to be of similar magnitude.

Therefore, a change from no IVH to grades 1 or 2 IVH should be similar in clinical importance to a change from grades 1 or 2 IVH to grades 3 or 4 IVH. An example of a non-randomised comparison (non-ITT) occurs with the outcome PDA closure. On day 1, for those neonates with serial echocardiography, PDA at any time on the first postnatal day was reported for 158/195 (81%) in the indomethacin group vs 187/206 (91%) in the placebo group. Such a chance imbalance is possible, but

the subsequent comparison at Day 5 is not technically by ITT, so not necessarily comparing like with like.

What can we learn from this trial?

This trial was investigating an important clinical question in a very sick patient population. It was well conducted but perhaps the analysis could have been pre-specified and therefore more transparent. In fairness, this trial was published before the first CONSORT statement for improving the reporting of randomised controlled trials (RCTs) was published.⁴

- iii. Using a final example of a trial published in 2001 studying the long-term effects of indomethacin prophylaxis in extremely-low-birth-weight (ELBW) infants.⁵

Prophylactic administration of indomethacin had been shown to reduce the frequency of PDA and severe IVH in very-low-birth-weight (VLBW) infants (<1,500g). The question remained however, whether prophylaxis with indomethacin conferred any long-term benefits that outweighed the risks of drug-induced reductions in renal, intestinal, and cerebral blood flow.

The design was a 'large' multi-centre placebo-controlled RCT with long term follow-up at 18 months age (corrected for prematurity). It built on the evidence base and looked at long term effects of indomethacin prophylaxis treatment in ELBW infants. The participants (**P**) were 1,202 new-born infants with birth weight 500–999g (considered for enrolment at 2 hrs old). The intervention (**I**) was indomethacin (0.1 mg per kilogram of body weight) or (**C**) saline placebo intravenously once daily for three days.

The primary outcome (**O**) comprised a composite of death, cerebral palsy, cognitive delay, deafness, and blindness at 18 months of age corrected for prematurity. Secondary short-term outcomes included closure of the PDA, pulmonary haemorrhage, chronic lung disease (CLD), ultrasound evidence of intracranial abnormalities, necrotising enterocolitis (NEC), and retinopathy of prematurity (ROP).

Out of 574 infants (with data on the primary outcome) assigned to prophylactic indomethacin, 271 (47%) died or survived with impairments, compared with 261 out of 569 infants (46%) assigned to placebo (odds ratio [OR], 1.1; 95% confidence interval [CI], 0.8 to 1.4; $P=0.61$). Selected outcomes illustrated in the previous examples were both significantly improved: indomethacin reduced the incidence of PDA (24% in the indomethacin group vs. 50% in the placebo group; OR 0.3 $P<0.001$) and severe periventricular and IVH (9% in the indomethacin group vs. 13% in the placebo group; OR 0.6; $P=0.02$). The authors concluded “In extremely-low-birth-weight infants, prophylaxis with indomethacin does not improve the rate of survival without neurosensory impairment at 18 months, despite the fact that it reduces the frequency of patent ductus arteriosus and severe periventricular and intraventricular haemorrhage”. Appropriately, the take home message was that “indomethacin prophylaxis should not be prescribed with the expectation that the chances of survival without neurosensory impairment will be improved”.

What were the strengths and weaknesses of this trial?

Strengths included the elegance of the design, clarity of reporting and the fact that the trial addressed an important clinical question. The internal validity (how well an experiment is done) was pretty exceptional: 92% of infants received ≥ 2 doses of trial medications (86% received their first dose within 6 hours), only 13/1,202 (0.01%) children were lost to follow-up, and adequate data for the analysis of the primary outcome (at a corrected age of 18 months) were available for 95% of infants. Other strengths included blinded assessment of outcomes, clear outcome definitions and a pre-specified ITT (adjusted) analysis.

What can we learn from this trial?

No trial is perfect though. Post hoc sample size/power calculations are not advised and one could debate whether the trial was large enough.⁶ However, one could argue that if a trial of 1,200 ELBW infants is not large enough to show benefit (or harm), then presumably the minimum clinically important difference threshold has not been reached. However, the main weakness of this and the

vast majority of trials investigating treatments for PDA is rescue treatment with open label treatment, in this case, indomethacin.

Of the 601 infants assigned to prophylaxis with indomethacin, 100 (17%) received additional indomethacin, compared to 276 of the 601 infants (46%) assigned to placebo (odds ratio, 0.2; 95% CI, 0.2 to 0.3; $P<0.001$).

Where next? Do we need evolution or revolution?

Questions remain regarding the use of drug treatment for an open PDA – why is that?

- i. First do no harm – some clinicians are unwilling to give a prophylactic treatment such as a drug to treat an open PDA given that a substantial proportion of those ducts will close without intervention (inversely proportional to gestational age), and all drugs carry with them short and long term side effects. The proportion of PDAs reported to be spontaneously closing ranges from 1 in 4 to as many as 4 in 5.⁷
- ii. Surgical ligation - is this option still pertinent to modern day RCTs?
- iii. Not all RCTs change clinical practice – fact, most are far too small to do so. Those that do change clinical practice are time-limited; some others mislead.
- iv. Use of surrogate outcomes – PDA closure was the first widely reported surrogate outcome of interest, however, one could debate whether this has delayed progress by distraction.
- v. RCTs compare group averages – large trials with broad eligibility criteria allow robust inference about how to treat the next ‘average’ patient. If the trials are large enough, they may even shed some light on subgroups that may benefit more than others. However, ostensibly, an RCT is a blunt instrument, not necessarily capable of precision targeted medicine.

Other potential reasons for the continued uncertainty

Misuse of post-hoc &/or subgroup analysis can perpetuate myths which then need debunking, again potentially delayed progress by distraction. This was previously widespread across medical research but should be reducing through initiatives such as education, trial registration, publication of protocols (and statistical analysis plans), the CONSORT statement and various extensions.

Technological &/or medical advances – one cannot future proof a trial's results, and technological advances such as echocardiography and ultrasound have allowed clinicians to diagnose, identify and target higher risk groups which may potentially benefit from treatment(s). It is plausible that blanket treatment yields little overall benefit, whilst beneath that there may be high risk subpopulation who could benefit (identified by improved diagnosis/risk stratification).

Equipoise issues – should one just treat, or rather take part in research to evaluate the best interventions in the absence of convincing evidence of treatment benefit, given that there is evidence that the presence of a PDA at 72 hours, in infants born less than 32 weeks of gestation, significantly increases the odds of death or severe morbidity compared to control infants without a PDA, especially in those infants with a large PDA ($\geq 1.5\text{mm}$)? ⁸ This dilemma has evolved, in part, from the clinical trials that have taken place in last three decades. The research landscape includes PDA trials studying babies up to 33 weeks of gestation, sometimes using surrogate outcomes and often with high open label treatment rates especially in the placebo arm against a backdrop of global improvements in perinatal care. In addition, other interventions for comorbidities are constantly evolving and alternative drugs/therapies emerge. Yesterday's findings may not be valid today and the today's question will change as perinatal care improves – a potentially fertile ground for more RCTs.

Applicability of the findings – changes in patient population. More infants are surviving at earlier gestations; babies are now surviving at 23-24 weeks of gestation. ⁹ This increase in survival at a given gestation with long term outcomes at 2 years of age was elegantly compared between two time

cohorts.¹⁰ The survival of extreme preterm babies has also been reported in EPICURE studies 10 years apart with an increase in survival but without changes in morbidity.¹¹

Methodological challenges to overcome

- i. *Open label (rescue) treatment* resulting in significant ‘contamination’ in both treated and placebo control groups – this is the greatest single threat to the reliability of the results. It skews the estimated treatment effect towards the null hypothesis of no difference between the groups. The level of contamination in RCTs varies greatly, ranging from 26% to 85%.¹²
- ii. *Intention-to-treat (ITT) analysis* – a consequence of the previous point. Most placebo-controlled randomised trials of PDA treatments are blinded, and should be; it is technically possible and scientifically the strongest approach to minimise performance and outcome ascertainment biases. Despite this, significant contamination of allocated groups occurs in such trials i.e. infants treated with drug are retreated and placebo infants also receive active treatment. This affects the treatment comparison that the trial was set up to examine by dilution. A bias is therefore imparted and some would suggest an ITT analysis is sub-optimal and advocate another form of analysis to be appropriate, such as the Complier-Average Causal Effect (CACE) analysis.¹³ However, other methodologists may disagree with this particular approach, but instead they would agree that avoiding the problem is the best solution. In addition, this ‘over’ treatment could have safety implications, and also could affect the integrity of the original research question. If clinicians are not in equipoise then they should refrain from enrolling an infant rather than enrolling an infant and then not following the protocol.
- iii. *Sample size* – bigger is better. Small trials can be misleading - the medical literature is littered with examples. The International Neonatal Immunotherapy Study (INIS) completely overpowered the hitherto gathered evidence on Intravenous immunoglobulin for suspected or proven infection in neonates.¹⁴ The subsequent Cochrane Authors’ conclusions were notably strong: “The undisputable results of the INIS trial, which enrolled 3,493 infants, and

our meta-analyses (n=3,973) showed no reduction in mortality during hospital stay, or death or major disability at two years of age in infants with suspected or proven infection.

Although based on a small sample size (n=266), this update provides additional evidence that IgM-enriched IVIG [intravenous immunoglobulin] does not significantly reduce mortality during hospital stay in infants with suspected infection. Routine administration of IVIG or IgM-enriched IVIG to prevent mortality in infants with suspected or proven neonatal infection is not recommended. No further research is recommended.”¹⁵

- iv. *Choice of outcomes* – surrogate vs substantive vs composite outcomes. Of course, the choice of outcome somewhat depends on the stage of the trial and how early it is on the pathway from bench to bedside. Potential intervention candidates need to show promising results on surrogates in order to go forward into the next phase of evaluation. So there clearly is a rightful place for surrogate outcomes. However, to change clinical practice one typically needs to show short &/or long-term benefits (clinical and cost-effectiveness) on substantive outcomes and most importantly, safety. Substantive outcomes are typically uncommon and therefore sample sizes are driven up, as are corresponding costs. Composite outcomes are one answer; although while they solve one problem they cause others. Components are typically not all as important as each other and the least important but more frequent ones can dominate. Interpretation of results can be problematic.¹⁶
- v. *Mortality* – death may not be the primary focus of the trial; indeed the intervention may have no plausible mechanism to affect mortality. But infant mortality in the higher risk populations can be high and therefore death cannot be ignored in the trial design, choice of outcomes or analysis.¹⁷
- vi. *Choice of assessment tools* – gold standard or fool’s gold? Trialists are limited in their arsenal of assessment tools, both in scope, their ability to detect subtle changes, and the ideal timing. One way forward is potentially to make greater use of routinely collected data gathered in, for example, the education system.

- vii. *More targeted patient selection and research questions* – the narrower the focus is in terms of patient subpopulations, the less applicable and potentially attractive to funders and clinical colleagues. Conversely, a big bang in terms of avoiding, for example, neurological impairment, could well be cost-effective over a lifetime and therefore very attractive to health care systems.
- viii. The issue of multiplicity and the incorrect analysis and misinterpretation of subgroup analyses presents a particular challenge – trials typically lack statistical power to reliably detect subgroup effects. In addition, there are numerous possible subgroups that can be examined; we need to guard against data dredging, which can lead to false claims. Finally, statistical significance (or lack of) in a specific subgroup is not the correct way to reinforce (or refute) any subgroup claims; instead, the statistical test of interaction should be used to directly infer whether the treatment effect appears to differ across subgroups.¹⁸
- ix. *Superiority vs non-inferiority designs* – it is not uncommon in the medical literature to see a lack of statistically significant evidence of superiority simply interpreted as equivalence. These designs are fundamentally different and should be treated accordingly.¹⁹
- x. *Consent* – (a) in a PDA trial, should one consent prior to ultrasound or after? The former could inadvertently result in a form of selection bias whereby a researcher decides who to approach knowing that they are eligible, as opposed to approaching all-comers to consent to being assessed for eligibility (the normal approach). The former mirrors clinical practice whereas the latter is more in keeping with research practice; (b) ‘Opt out’ consent – in order to increase the number of participants recruited into trials and research capacity per se (thereby making trials finish sooner), an opt-out consent culture would be ideal. It would help ‘normalise’ research in the eyes of the research community and parents too.

How should trials for treating an open PDA look in 2020 and beyond – do we need evolution or revolution?

Evolution might involve more two-arm parallel-group (individually randomised or cluster randomised) RCTs comparing a new intervention vs the standard of care ('treatment as usual'). This is the most simple trial design which answers one question reliably, but is it cost-effective to continually perform a series of two-arm RCTs?

An alternative would be greater use of multi-arm RCTs &/or care pathway trials. The latter would be very attractive to funders such as the National Institute for Health Research (NIHR) in the UK. An innovative trial design like that published by Gersony *et al*, would mirror present clinical practice and thereby have a higher chance of success, whilst potentially answering more than one research question.

A more conventional approach would be using a multi-arm RCT – for example, the three arm 'gold standard' design. For a condition where a commonly acknowledged reference standard drug therapy exists, if it can be justified on ethical grounds, one might recommend evaluating the efficacy and safety of a new intervention in a three-arm trial design with the reference standard, placebo and the investigational drug.²⁰

The trial could aim to show:

1. superiority of the investigational drug over placebo (proof of efficacy)
2. superiority of the reference standard over placebo (proof of assay sensitivity)
3. the investigational drug retains most of the efficacy of the reference standard (proof of non-inferiority)

Revolution might involve a Multi-Arm, Multi-Stage platform trial (MAMS).

MAMS platform trials have two elements. The multi-arm element facilitates several candidate treatments to be simultaneously assessed against a common control group, within a single

randomised trial. Whilst in the multi-stage element, participant recruitment is discontinued to treatment groups that are not showing sufficient promise based on a series of pre-specified, interim, lack-of-benefit analyses.²¹

MAMS in context of PDA treatment evaluation could entail an early phase (randomised phase II) trial using surrogate outcomes moving seamlessly into later phase (III) using substantive outcomes. For example, the first element evaluates 3 treatments X vs Y vs Z on the outcome PDA closure, leading to a 2 treatment head-to-head trial of X vs Z on the outcome 2 year neurological development. Other candidate treatments can be added in the future to be compared with the reference standard.

Summary – what can we learn from the past?

History teaches us important lessons. Some of the greatest advances were made a long time ago e.g. R Gross performed the 1st successful ligation of a PDA on 8 year-old girl at Children's Hospital Boston in 1938. That same year RA Fisher stated “To consult the statistician after an experiment is finished is often merely to ask him to conduct a post-mortem examination. He can perhaps say what the experiment died of.” It was another 34 years before the first reports of successful operative closure of PDA in premature babies were published in 1972. Gersony et al. published their ‘care pathway’ trial in 1983, the year that SJ Pocock published the first edition of his text book “Clinical Trials: A Practical Approach”. It was another 18 years before Schmidt et al. published their trial on indomethacin prophylaxis in 2001, the year Moher et al. published the CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials.

As a research community we need to learn from the past and agree on a way forward to get to the answers reliably and quickly. Conferences like NeoCARD 2017, bringing together experts and stakeholders from around the globe, help to formulate opinion and knowledge, and hone, in this case, the PDA research agenda to the benefit of everybody.

Clinical Practice Points

1. The majority of clinical trials in the past 50 years have evaluated prophylactic treatments for the closure of a PDA. Some included babies up to 34 weeks of gestation as opposed to today's PDA management dilemma in extreme preterm babies.
2. Biomarkers are now available bedside that facilitate the early detection of a 'haemodynamically' significant PDA and so new approaches can be evaluated for the management of PDA.

Research Directions

Well-designed clinical trials can provide robust evidence which can ultimately change clinical practice. Trial methodologists should be involved from the outset and great care should be taken to design trials which:

1. Minimise 'contamination' (i.e. dilution of the treatment effect) of the allocated groups, and thereby preserve the integrity of the intention to treat analysis by keeping the level of open label 'rescue' treatment by clinicians to the bare minimum, without endangering participants (linked to equipoise);
2. Use a large enough sample size to be convincing (with the exception of pilot/feasibility studies);
3. Consider alternative as well as conventional designs, especially the 'care pathway' model;
4. Avoid, where practicable, the use of surrogate outcomes.

Acknowledgements

The speakers at NeoCARD 2017.

Karen Juszczak for her help in refining the themes within the conference presentation that preceded this article.

Ann Kennedy for her help with the history timeline within the conference presentation that preceded this article, and Jenny Shilton Osborne for her help proof-reading, commenting upon and re-formatting the article.

Health Care Professionals all over the UK.

Most importantly of all, parents who enrol their babies in trials run by the NPEU Clinical Trials Unit, University of Oxford <https://www.npeu.ox.ac.uk/trials>

Legend

Figure 1. Schema of study design, summarizing randomization allocations on diagnosis of hemodynamically significant PDA in trial A and subsequent randomization in trial B of infants initially receiving placebo who required backup therapy.

Reprinted from J Pediatr. 1983 Jun;102(6): 895-906. Gersony WM, Peckham GJ, Ellison RC, Miettinen OS, Nadas AS. Effects of indomethacin in premature infants with patent ductus arteriosus: results of a national collaborative study. Copyright (1983), with permission from Elsevier.

References

¹ <http://www.jameslindlibrary.org/> and Wikipedia

² Gersony WM, Peckham GJ, Ellison RC, Miettinen OS, Nadas AS. Effects of indomethacin in premature infants with patent ductus arteriosus: results of a national collaborative study. *J Pediatr*. 1983 Jun;102(6):895-906.

³ Ment LR, Oh W, Ehrenkranz RA, Philip AG, Vohr B, Allan W, Duncan CC, Scott DT, Taylor KJ, Katz KH, et al. Low-dose indomethacin and prevention of intraventricular hemorrhage: a multicenter randomized trial. *Pediatrics*. 1994 Apr;93(4):543-50.

⁴ Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D, Stroup DF. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA*. 1996 Aug 28;276(8):637-9.

⁵ Schmidt B, Davis P, Moddemann D, Ohlsson A, Roberts RS, Saigal S, Solimano A, Vincer M, Wright LL; Trial of Indomethacin Prophylaxis in Preterms Investigators. Long-term effects of indomethacin prophylaxis in extremely-low-birth-weight infants. *N Engl J Med*. 2001 Jun 28;344(26):1966-72.

⁶ Goodman SN, Berlin JA. The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. *Ann Intern Med*. 1994 Aug 1;121(3):200-6. Erratum in: *Ann Intern Med* 1995 Mar 15;122(6):478.

⁷ Benitz WE; Committee on Fetus and Newborn, American Academy of Pediatrics. Patent Ductus Arteriosus in Preterm Infants. *Pediatrics*. 2016;137(1):e20153730

⁸ Sellmer A, Bjerre JV, Schmidt MR, McNamara PJ, Hjortdal VE, Høst B, Bech BH, Henriksen TB. Morbidity and mortality in preterm neonates with patent ductus arteriosus on day 3. *Arch Dis Child Fetal Neonatal Ed*. 2013 Nov;98(6):F505-10.

⁹ Patel RM. Short- and Long-Term Outcomes for Extremely Preterm Infants. *Am J Perinatol*. 2016 Feb;33(3):318-28.

¹⁰ Tin W, Wariyar U, Hey E. Changing prognosis for babies of less than 28 weeks' gestation in the north of England between 1983 and 1994. Northern Neonatal Network. *BMJ*. 1997 Jan 11;314(7074):107-11.

-
- ¹¹ Costeloe KL, Hennessy EM, Haider S, Stacey F, Marlow N, Draper ES. Short term outcomes after extreme preterm birth in England: comparison of two birth cohorts in 1995 and 2006 (the EPICure studies). *BMJ*. 2012 Dec 4;345:e7976.
- ¹² Evans N. Preterm patent ductus arteriosus: A continuing conundrum for the neonatologist? *Semin Fetal Neonatal Med*. 2015 Aug; 20(4):272-7.
- ¹³ Ye C, Beyene J, Browne G, Thabane L. Estimating treatment effects in randomised controlled trials with non-compliance: a simulation study. *BMJ Open*. 2014 Jun 17;4(6):e005362.
- ¹⁴ INIS Collaborative Group, Brocklehurst P, Farrell B, King A, Juszczak E, Darlow B, Haque K, Salt A, Stenson B, Tarnow-Mordi W. Treatment of neonatal sepsis with intravenous immune globulin. *N Engl J Med*. 2011 Sep 29;365(13):1201-11.
- ¹⁵ Ohlsson A, Lacy JB. Intravenous immunoglobulin for suspected or proven infection in neonates. *Cochrane Database of Systematic Reviews* 2015, Issue 3. Art. No.: CD001239. DOI: 10.1002/14651858.CD001239.pub5.
- ¹⁶ Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C. Composite outcomes in randomized trials: greater precision but with greater uncertainty? *JAMA*. 2003 May 21;289(19):2554-9.
- ¹⁷ Parekh SA, Field DJ, Johnson S, Juszczak E. Accounting for deaths in neonatal trials: is there a correct approach? *Arch Dis Child Fetal Neonatal Ed*. 2015 May;100(3):F193-7.
- ¹⁸ Pocock SJ, McMurray JJV, Collier TJ. Statistical Controversies in Reporting of Clinical Trials: Part 2 of a 4-Part Series on Statistics for Clinical Trials. *J Am Coll Cardiol*. 2015 Dec 15;66(23):2648-2662.
- ¹⁹ Pocock SJ, Clayton TC, Stone GW. Challenging Issues in Clinical Trial Design: Part 4 of a 4-Part Series on Statistics for Clinical Trials. *J Am Coll Cardiol*. 2015 Dec 29;66(25):2886-98.
- ²⁰ EMEA/CPMP/EWP/908/99 Points to Consider on Multiplicity Issues in Clinical Trials (Sept 2002)
- ²¹ http://www.ctu.mrc.ac.uk/our_research/research_types/methodology1/design/mams_design/