

The challenges of randomised control trials in obstetrics and gynaecology

Sarah Harper MRCOG Consultant Obstetrician and Gynaecologist, United Lincolnshire Hospitals NHS Trust, Lincoln, LN2 5QY, UK.

sarah.harper@ulh.nhs.uk

Jim G Thornton FRCOG, MD Professor of Obstetrics and Gynaecology, Division of Child Health, Obstetrics and Gynaecology, School of Medicine, University of Nottingham, Maternity Department, Nottingham City Hospital, Nottingham University Hospitals NHS Trust, Nottingham, NG5 1PB UK.

jim.thornton@nottingham.ac.uk

Kate F Walker MRCOG, PhD Clinical Associate Professor in Obstetrics, Division of Child Health, Obstetrics and Gynaecology, School of Medicine, University of Nottingham, Maternity Department, Queen's Medical Centre, Nottingham University Hospitals NHS Trust, Nottingham, NG7 2UH, UK.

kate.walker@nottingham.ac.uk

Abstract and key words

In this short article we explore some of the challenges faced by those conducting randomised trials in obstetrics and gynaecology. We discuss the current status of trials comparing induction of labour versus expectant management, so called 'deliver or delay' trials. We consider the benefits of utilising routine data for collecting trial outcome data. Although cluster trials can provide a useful methodology for answering difficult questions, we illustrate with an example that cluster trials are at risk of delivering a misleading result. We discuss the importance of long term outcomes in randomised trials.

Key words: randomised controlled trial; cluster randomised trial; routine data; long term outcomes

Background

Randomised controlled trials are regarded as the gold standard study design to determine the effectiveness of a new treatment. Randomising patients to different treatment arms aims to eliminate selection bias. Selection bias occurs when a recruiter selectively enrolls patients to a study based on advanced knowledge of what the study arm they will be allocated to. This leads to an imbalance in the characteristics of the participants in each arm of the study. If randomisation is performed correctly, it eliminates selection bias and leads to a balance of patients with different characteristics (e.g. BMI, age, and ethnicity) in the study arms. This can then make us confident that any observed difference in the effects of the interventions

in the study arms is due to the effects of the intervention itself not simply due to the intervention being tested in a particular group of participants.

Deliver or delay trials

The question of whether to intervene and recommend delivery or 'manage expectantly' i.e. waiting until a situation arises that necessitates recommending delivery (so called 'deliver or delay' trials) has culminated in over 30 trials and three systematic reviews. Recently the US ARRIVE trial randomised more than 6000 low risk women with no indication for induction and showed that induction of labour 39⁺⁰ to 39⁺⁴ for low risk women conferred no benefit to the baby. Women in the induction arm had a reduced risk of caesarean and pre-eclampsia, but no difference was observed in the composite adverse perinatal outcome including stillbirth. We must be mindful that the findings of ARRIVE may not apply to a UK population with a very different health care system and population.

The abundance of evidence of the benefits of induction from 39 weeks gestation onwards for various high risk women has contributed to a national induction of labour rate of 31%, this has a significant impact on patient experience (due to delays in the induction process) and work load for staff.

Routine data

The maternity services data set (MSDS) provides patient level data from booking to discharge from maternity services on the mother and baby. MSDS data is then reported to and held by NHS Digital. Trials can now be performed utilising routine data from national sources, with the correct permissions, eliminating the need for individual data to be collected on a participant by participant basis, this makes previously considered impossible trials due to their size and subsequent cost, now possible.

The GBS3 trial (ISRCTN49639731) is a cluster-randomised trial which will determine the clinical and cost-effectiveness of routine testing in late pregnancy or labour for Group B streptococcus, compared to the current risk-factor based strategy. It requires 320,000 women to have adequate power to show a difference between a risk factor approach and routine testing as neonatal GBS infections are so rare. Using routine data that is already collected on the MSDS avoids overburdening busy clinical staff and avoids laborious painstaking data collection by research teams and its associated costs.

The AFFIRM trial utilised routine data. The AFFIRM trial was a stepped-wedge, cluster-randomised trial which assessed a programme that encouraged enhanced maternal awareness and rapid reporting of changes in fetal movement, which was combined with training of staff to respond with a defined programme of further testing and, if necessary, to induce delivery. Control centres gave usual care, and the primary outcome was stillbirth. More than 400 000 pregnancies at 33 hospitals were included, so the study was powered to exclude even modest effects. The intervention was associated with an induction rate of 41%, compared with 36% in the control group (adjusted odds ratio [AOR] 1.05, 95% CI 1.02–1.08); if the correct women had been induced and no harm caused, this intervention should have had a substantial effect. Unfortunately, stillbirths were not significantly reduced by this

intervention (AOR 0.90, 95% CI 0.75–1.07) and there was no effect on perinatal mortality (0.98, 0.83–1.17).

Although reduced fetal movement is now one of the most common indications for induction this study showed that a 'reduced fetal movement package' did not reduce the risk of stillbirth. Saving Babies Lives care bundle employed multiple interventions such as smoking cessation and increased induction resulting in a reduction in stillbirth rates. The AFFIRM trial does not suggest that reporting and acting on reduced fetal movements has an impact on the stillbirth rate as a single action. Yet in practice we still emphasise the importance of fetal movements at each contact and place significant pressure on the woman to report concerns. The use of routine data allowed this huge, well designed trial to be performed and our interpretation is that induction of labour for reduced fetal movements in the presence of normal fetal surveillance should be avoided less than 39 weeks gestation.

Cluster trials

A word of caution with cluster trials. Cluster trials are at very high risk of giving a misleading result. A stepped wedge cluster trial (PARROT trial) of placental growth factor (PIGF) as a diagnostic tool for women with suspected pre-eclampsia reported that PIGF testing leads to earlier diagnosis, reduces maternal adverse events (4% v 5%, AOR 0.32, 95%CI 0.11–0.96; $p=0.043$) but not fetal adverse outcomes (15% v 14%, AOR 1.45, 95%CI 0.73–2.90). We believed this is weak justification for implementation especially prior to independent judgment of clinical and cost-effectiveness. Although not statistically significant, the mean two day shorter gestation and 63gm weight difference suggests overall births were earlier. Ending pregnancy always improves maternal but not necessarily child outcomes.

More importantly "when the outcome [of a stepped wedge trial] requires individual participant data collection, lack of concealment of allocation is likely to mean a risk of differential selection of participants between arms." This affected PARROT. Although the same units acted as control and intervention centres, with identical recruitment periods, there were 127 (28%) more participants during intervention. 'Recruitment bias' means these are likely to have had milder disease and fewer complications. The authors optimistically suggest that "these biases do not appear to be substantial" but cannot know this.

Crudely modelling what happens if all additional recruits had no adverse events, by inflating the control denominator, abolishes the overall reduction in adverse maternal outcome (4% both groups), and increases adverse baby outcomes (15% v 11%) favouring controls. We are not alone in these conclusions.

Long term outcomes

As obstetricians we often assume wrongly our job is done once the baby is born, but we have an obligation to the baby and mother to do no harm, not just in the short term, but also the long term. Until recently little emphasis had been placed on the long term outcome of the "early term" babies we deliver between 37-39 weeks. A systematic review which included 41,344 children found increased cognitive scores for infants born at full term (> 39weeks) compared to those born at early term 37-38 weeks. Delivery at 37 weeks cannot be considered without an exploration of risks to the infant and should only be undertaken when the risks of continuing with the pregnancy outweigh the risks including in the long term to the

infant. These findings should be discussed with the mother to allow her to make an informed decision about induction of labour before 39 weeks.

The landmark ORACLE trial showed neonatal benefits from maternal administration of antibiotics for preterm ruptured membranes and this widely changed practice. The follow up study of infants saw multi – attribute health scores and Key Stage 1 attainment collected via questionnaires and SATS results. In the preterm rupture of membranes group there was insignificant effect on health and educational attainment. In the spontaneous labour group there was an increased risk of cerebral palsy and functional impairment associated with antibiotics. This was increased further where women had received both erythromycin and co-amoxiclav. The absence of benefit is not unexpected but harm is surprising.

The ORACLE studies lend weight to the need to consider long term outcomes for the infant. Trials need to be adequately funded from the outset to look at longer and more comprehensive outcomes. Neurodevelopmental function in the long term should be considered a primary outcome. Long term outcomes should be standard and where this is not considered justification should be offered.

The current landscape

All UK non COVID-19 prioritised research was paused during the COVID-19 pandemic, this prevented recruitment into already established RCTs and the set-up of new trials in O&G and all other specialties. Some health services were reduced during this time in the face of rising infections, but this was simply not possible in maternity care. Maternity services have carried on and continue to do so throughout the pandemic which now with a 2nd wave in full force shows no signs of relenting. Given the necessary continuation of maternity care many trials which could have continued given the ongoing services provided were paused in our view unnecessarily. It is not in the interest of women and babies for research to be put on hold in a blanket fashion, a more nuanced approach is necessary. The National Institute for Health Research have now issued guidance that non-COVID research should be protected in the second wave of infections.

Conflicts of interest

None to declare

Role of the funding source

No funding received

Key references

1. Grobman, W.A., Labor Induction vs. Expectant Management of Low-Risk Pregnancy. *N Engl J Med*, 2018. 379(23): p. 2278-2279.
2. Norman, J.E., et al., Awareness of fetal movements and care package to reduce fetal mortality (AFFIRM): a stepped wedge, cluster-randomised trial. *Lancet*, 2018. 392(10158): p. 1629-1638.
3. Duhig, K.E., et al., Placental growth factor testing to assess women with suspected pre-eclampsia: a multicentre, pragmatic, stepped-wedge cluster-randomised controlled trial. *Lancet*, 2019. 393(10183): p. 1807-1818.
4. Hemming, K., et al., The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. *BMJ*, 2015. 350: p. h391.
5. Tanner, M. and B.W. Mol, Placental growth factor testing in suspected pre-eclampsia. *Lancet*, 2020. 395(10221): p. 335.

Self-Assessment Question

SBA

Randomised controlled trials are considered the gold standard study design for evaluating new interventions because

- A They can take the form of a cluster trial
- B The investigator is able to allocate a participant to the treatment arm that suits the participant best
- C Randomising individuals to the different treatment arms in the study eliminates selection bias
- D They are not affected by any types of bias
- E Randomising individuals to the different treatment arms in the study leads to selection bias