Summary of Discussion

Boulvain M, Senat MV, Perrotin F, et al. Induction of labour versus expectant management for large-for-date fetuses: a randomised controlled trial. Lancet. 2015;385(9987):2600-5

John focused conversation around the "super seven" issues

- What is the PICO (Population, Intervention, Control, Outcome) of this paper?
- How did the researchers randomise the groups of patients, and are we happy with their approach?
- Are the groups well matched?
- Is blinding possible? If not, why not?
- Does the study have enough patients recruited to show that any effectiveness shown was not achieved by chance?
- What are the results?
- Can we trust the results, are they statistically significant?

The notes below were shared with delegates after the session

∙∙

We agreed that this is an RCT which is considered to be the gold standard in measuring the effectiveness of a drug or any intervention.   There was no blinding which often appears to reduce the apparent effectiveness of the intervention, but we discussed the fact that this was a 'person-centred' trial – and how that aspect was unavoidable.

It works with the concept of a control and an intervention (the intervention is usually a drug and the control is often a placebo that looks like the drug) we then compare the outcomes to get a true view of the effectiveness of the intervention.

WHAT IS THE PICO?

We discussed the PICO (Population Intervention, Comparator, Outcome). This helps focus the mind on the key aspects of an EFFECTIVENESS trial (effectiveness covers RCT, Systematic Review and Cohort studies)

P   Women with singleton fetuses whose estimated weight exceeded the 95th percentile

I   Induction of labour within 3 days between $37+^0$ weeks and $38+^6$ weeks of gestation,

C   Expectant Management

O   Primary outcome was a composite of clinically significant shoulder dystocia, fracture of the clavicle, brachial plexus injury, intracranial haemorrhage, or death

1)   Did the study ask a clearly focussed question?

Yes – does induction of labour within 37 and 38 weeks of gestation reduce a composite of negative outcomes (shoulder dystocia, clavicle fracture, brachial Plexus Injury, intracranial Haemorrhage or death?

2)   Was this a randomised Controlled Trial (RCT) and was it appropriately so?

Yes – again because it is an effectiveness question (see above) - see randomisation on page 2601 – they used computer- generated permuted block with a block size of 4 to 8.   They stratified by centre.

Why was this study carried out as an RCT?

Because there needed to be some comparison of groups in similar situations it was a pragmatic randomised placebo controlled trial. Clinicians were not masked to group allocation and patients clearly would know which group they were in.

Is it worth continuing?

3)   a) Were participants appropriately allocated to intervention and control groups?

How patients were allocated to intervention and control groups – was the process truly random?

Baseline characteristics - there is a very balanced randomization (see table 1 page 2602), age, previous history of macrosomia broadly similar. Expectant management group had a higher weight gain in pregnancy, but the induction of labour group were heavier before pregnancy.

b)   Was a methods used to balance the randomization (eg stratification)

Stratification?   where prognostic variables have to be allowed for in the trial allocation.   External differences in the group need to be catered for – e.g. breast cancer – you might want to stratify a trial to get an even number of post menopausal women on each side.   If you don't do this, you may unintentionally be comparing patients with different prognosis and that would not be a credible trial.   A lot of people have trouble with stratification because they think that it is

tinkering with the random nature of a trial

Yes it was stratified for centre (i.e hospital (page 2) – we wondered whether it should it have been stratified for previous pregnancy

c)   Are there any differences reported that might have explained the outcome (confounding)?

No obvious confounders

4)   Were participants, staff and study personnel 'blind' to participants study group?

No – patients and clinicians were not blinded – the only way this could have been blinded was to have had researchers collating the negative outcomes while unaware of which group the patients were in.

5)   Were all of the participants who entered the trial accounted for at its conclusion?

See page 2601 – Trial Profile.   Only 4 patients were lost to follow up and that was before delivery, so this will not have affected outcome measurement [unless they were frightened of induction or going too long – there was a slight concern that fear of having a large neonate might encourage them to want induction.

6)   Were the participants in all groups followed up and data collected in the same way – yes page 2 – the final follow up was fixed at 4 weeks

7)   Did the study have enough participants to minimise the play of chance [POWER]?

They based the initial sample size calculation on detection of a difference in percentages of the primary outcome, with a power of 80% and a type 1 error of 5%.

They assumed the risk in the control group to be 5–10% and the risk in the induction of labour group to be 1·65–5% (ie, an RR of 0·33–0·50). The calculation showed that a total sample size of about 1000 women (500 per group) was sufficient to show these differences. They didn't get enough patients due to financial constraints which made it necessary to end recruitment at a predetermined date (Jan 1, 2009), before they did any analyses.   We had some concerns about this as it appeared to reduce the effectiveness data.

Results – We calculated how large the treatment effect was

Outcome: There were **no** brachial plexus injuries, intracranial haemorrhages, or perinatal deaths.   Induction of labour significantly reduced the risk of shoulder dystocia or associated morbidity – 8 in the induction group, 25 in the expectant management. They quote absolute risk as 4% and NNT as 25

We checked that and calculated the raw data.

We know that there were 8 cases of the combined outcome in the induction group and 25 in the control (Expectant management) group (see table 2 on page 2602)

Outcome: Shoulder Dystocia

	Yes	No	
Experimental group (Induction)	8	399	407
Control group (Expectant management)	25	386	411
	33	785	818

1. How many people are in the study?	818
2. How many people were allocated to the control group?	411
3. How many people were allocated to the experimental group?	407
4. How many people in the experimental group had the combined outcome?	8
5. How many people in the control group had the combined outcome?	25
6. How many people had the combined outcome altogether?	33
7. How many people in the experimental group did not have the combined outcome?	399
8. How many people in the control group did not have the combined outcome?	386

9. How many people did not have the combined outcome in total?	785
10. What are your chances of having the combined outcome in the experimental group? $\quad\quad\quad\quad$ a/(a+b) 8/407 = 1.96% *NB this is known as the **experimental event rate (EER)***	2%

1. What are your chances of having the combined outcome if you are in the control group? $\qquad$ c/(c+d) 25/411 = 6.08% *This is known as the **control event rate (CER)***	6%
12. How does the combined outcome rate in the experimental group compare with thecombined outcome in the control group? CER – EER *This is known as the **absolute risk reduction (ARR)***	6-2 4%
13. How many people would have to have induction in order to prevent 1 additional person having the combined outcome? *100/ARR 100/4* *NB This is known as the **number needed to treat (NNT)***	25
16. To work out 95% confidence intervals, use the following formula; $+/-1.96\sqrt{\dfrac{\text{CER(100-CER)}}{\text{n (in control)}} + \dfrac{\text{EER(100-EER)}}{\text{n (in experimental)}}}$ To find the 95% CIs around the ARR, add and subtract this number from the ARR. To find the 95% CIs for the NNT, divide the new limits for the ARR into 100.	Formula total = ∴ARR = 4% (1.4-6.8) NNT = 25 (15-71)

This means that you need to treat 25 people with induction to prevent 1 additional case.

They have used relative risk which looks more attractive, but they did calculate the risk difference as we did on page 2603.

7) How precise are the results?

They quote confidence intervals which show the uncertainty in any study. For example, the absolute risk reduction of 4% is given a confidence interval of 1.4& to 6.8% which is very wide considering the small difference – meaning it lacks precision. This is even worse in the confidence interval of the NNT 15-70

The good news is that they are not doing any harm as the lower end didn't cross the line of no effect.

So confidence interval is (ARR = 4) 1.3 – 6.7

NNT = 25 (confidence Interval 100/1.3 – 100/6.7

15-76 therefore lacking in precision

8)   Were all important outcomes considered?

        What other information do you require to make a decision?

        Cost in terms of bed time savings – better outcome for women and presumably cheaper?

John Blenkinsopp
January 2021