RANDOMISED CONTROLLED TRIALS – *ARE THEY ALWAYS THE GOLD STANDARD?*

A/PROF JANE WARLAND

stillaware.org/usa2

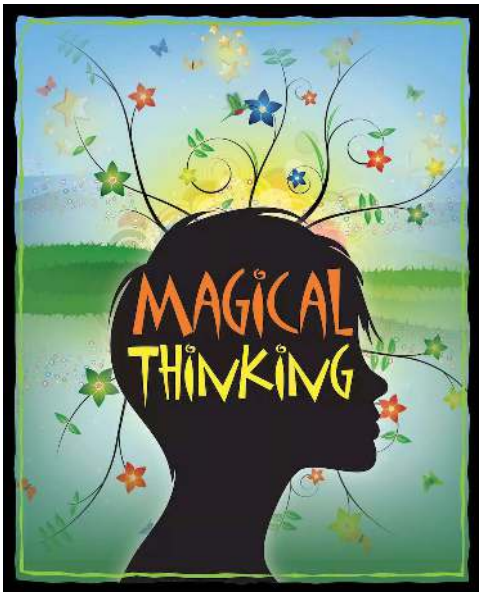
LEARNING OBJECTIVES

After attending this presentation attendees will:

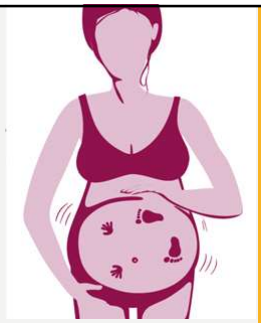
- Discuss findings from recent RCTs (particularly AFFIRM)
- Identify the advantages and disadvantages of conducting a randomised controlled trial when a primary outcome measure is rate of stillbirth
- Explain potential pitfalls for conducting studies to identify interventions to prevent (or reduce) stillbirth using "gold standard" methodologies

We are not against RCT's only magical thinking about them" (Deaton and Cartwright 2019)

..few things annoy us more than the deification that clinicians and selected researchers have given to randomize controlled trials (Cook and Thigpen 2019).



AFFIRM: WHAT IT WAS



- HYPOTHESIS: that rates of stillbirth will be reduced by introduction of a package of care consisting of strategies for increasing pregnant women's awareness of the need for prompt reporting of decreased fetal movements, followed by a management plan for identification of placental insufficiency with timely birth in confirmed cases.

AFFIRM: WHY THEY SAID THEY DID IT?

- Stillbirth dropped by 30% after the introduction of a similar package of care in Norway *but the efficacy of this intervention (and possible adverse effects and implications for service delivery) have not been tested in a randomised trial.*
- *BUT it was not randomised, and therefore constitutes only level II-3 evidence, it has led to new recommendations from the RCOG that "women should be advised to be aware of their baby's individual pattern of movements and that if they are concerned about a reduction in or cessation of fetal movements …..they should contact their maternity unit" …..*
- *In AFFIRM study we plan to formally test (using gold standard methodologies) whether a similar package of interventions really does decrease stillbirth, whether it does any harm (e.g. by increasing rates of caesarean section or induction of labour) and how it can be implemented to best effect in a very different setting* (Norman 2014) .

WHAT IS A STEP-WEDGE CLUSTER ?

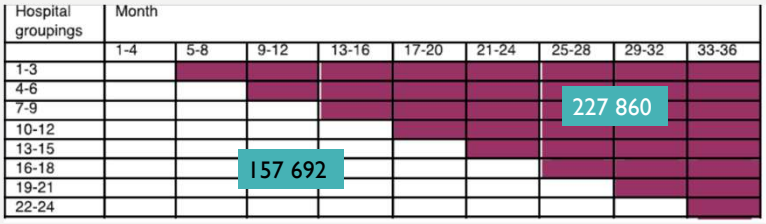
- Hospitals (not people) randomised to the timing of the introduction of an intervention
- All clusters in a stepped-wedge trial will receive the new intervention, the time at which they do so is determined by chance
- Used when randomisation of people to non-intervention is thought to be unethical or not feasible

Table 1: Stepped wedge design. Shaded areas indicate periods in which the interventions are being implemented. The order in which hospital groupings implement the interventions will be determined via randomization.

Hospital groupings	Month								
	1-4	5-8	9-12	13-16	17-20	21-24	25-28	29-32	33-36
1-3		■	■	■	■	■	■	■	■
4-6			■	■	■	■	■	■	■
7-9				■	■	■	■	■	■
10-12					■	■	■	■	■
13-15						■	■	■	■
16-18							■	■	■
19-21								■	■
22-24									■

AFFIRM RESULTS

- 33 hospitals were randomly assigned to an intervention implementation date.
- Data were collected from 409 175 pregnancies (157 692 births during the control period, and 227 860 births in the intervention period).
- The incidence of stillbirth was 4.40 per 1000 births during the control period and 4 ·06 per 1000 births in the intervention period (aOR 0 ·90, 95% CI 0 ·75–1 ·07; p=0.23).
- Induction of labour (aOR1 ·05 (1 ·02,1 ·08) p=0 ·0015) and caesarean section (aOR 1 ·09 (1 ·06–1 ·12) p<0 ·0001) were slightly more common during the intervention period than during the control period.



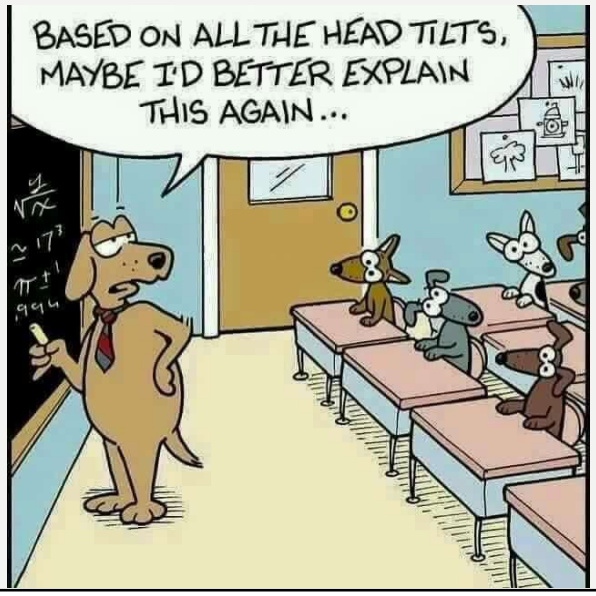
HEADLINES

Fetal Movement Awareness Flops for Stillbirth Prevention
-British-Irish study found no reduced risk
by Molly Walker, Staff Writer, MedPage Today September 28, 2018

READING BEYOND THE HEADLINES

- The title of the editorial 'encouraging awareness of fetal movement is harmful' does not accurately reflect the AFFIRM trial findings.
- It is important to look beyond the headlines and note:
  - Stillbirth reduced by 8.9% This effect, if confirmed in ongoing studies, could translate into over 4000 stillbirths alone averted annually (and families spared the tragedy of this loss) across high income countries
  - Awareness was not reported as being assessed.
  - The uptake of the AFFIRM intervention by clinicians was also not reported as having been assessed, i.e. so we do not know how well it was implemented.
  - Therefore current practices around awareness raising and clinical management around RFM should remain unchanged.

UNDERSTANDING THESE RESULTS



My Imaginary Study

Research question:
Does giving young children written information about the appearance of common household items increase their ability (awareness) to locate said items?

Approach:
Step wedge cluster RCT.
Sheds were located and randomised NOT the children

Methods:
A glossy brochure was produced and given to parents to give to the children. A link to a "how to talk to children about hammers!" e learning package was sent to all participating families.

Statistical power:
An earlier observational study showed that children can locate a hammer approximately 30% of the time, so this study expected to demonstrate at least that

Findings:
Only 9% of the children could locate a hammer. 28% got a splinter and 40 % got bitten by a red back spider

Conclusions:
Providing parents and children information about hammer apearance is an unproven strategy to raise awareness of what a hammer looks like and may do more harm than good

(the HAMMER trial: household awareness of manipulative material items for early readers.

APPROACH:

This sounds exciting?

Step wedge cluster RCT.
Sheds were located and randomised NOT the children

Comment:

The nature of the sheds (size, number of hammers in the shed (if any), the state of the shed all need to be assessed, as well as the possibility of a tidy up

METHOD

Did I hear you right? You don't like reading??

Comments about Methods:
A glossy brochure was produced and given to parents to give to the children

- How was the brochure given to the children?
- Was the brochure suitable for all age groups, reading ages and cultures
- Did the parents engage with the study or not?
- Did the child read it themselves, or not

Did the parents engage with the e-learning package?

FINDINGS

Findings:
Only 9% of the children could locate a hammer. 28% got a splinter and 40 % got bitten by a red back spider

Can you come with me?

COMMENTS ABOUT POWER

Power Calculations safeguard against:
The trial failing to detect something that is actually there by having enough participant numbers?

BUT

Sample size is a limitation since it can compromise the conclusions drawn from the studies. Too small a sample may prevent the findings from being extrapolated, whereas too large a sample may amplify the detection of differences, *emphasizing statistical differences that are not clinically relevant*. (Faber & Fonseca 2014)

Is this it?

The Hammer was there and she didn't recognise it
OR
It really isn't there so that's why she brought you something else

LOOKING FOR THE HAMMER

Splinters

Hammers

Spiders

LOOKING FOR THE HAMMER

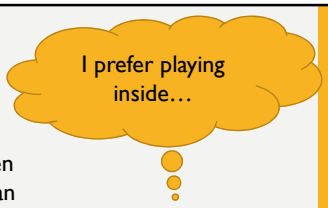
Splinters

Hammers

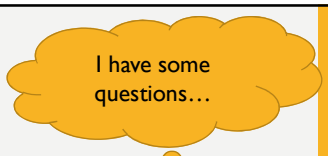
Spiders

CONCLUSIONS

Providing parents and children information about hammer appearance is an unproven strategy to raise awareness of what a hammer looks like and may do more harm than good



QUESTIONS



BUT is that a fair conclusion given we don't know……

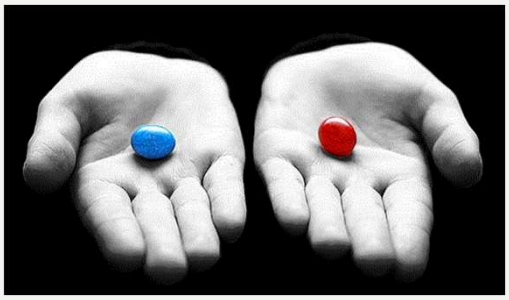
IF the brochure was given to the child
How the content of the brochure was communicated to the child
If the child understood what a hammer looked like
If the child already knew what a hammer looked like
If the parents didn't bother to give the brochure to the child because they thought the child already knew what a hammer looked like
The child had previous experience in locating a hammer

We also don't know
How long the child was encouraged to look in the shed
If their parents went with them
If the parents engaged with the e learning package
If the children who were 'harmed' were a subset of children i.e. the children who read the brochure on their own, went to the shed alone or children who like playing with spiders

Fair??...Reasonable??

WHEN IS RCT THE "GOLD STANDARD"

- The best way to compare a new treatment to the standard treatment is in a randomised controlled trial. In such a study, participants are randomly allocated to either the new or standard (control) treatments. This process is known to be an unbiased estimate of the treatment effect.



WHEN MIGHT A RCT NOT BE THE "GOLD" STANDARD ?

- "Intervention" is subjective and open to interpretation
- Equipoise
- Inadequate or inappropriate sample size calculation

THE AFFIRM INTERVENTION

Pamphlet to be given to pregnant women BUT there was

No mention in the manuscript about :

- If a standardised gestation, "*about 20 weeks*"
- If a script used
- If understanding was measured
- If awareness was measured
- How many care providers accessed the e-learning

the intervention package might not have been sufficiently effective to initiate behaviour change in clinicians and in pregnant women (Norman et al 2018).

INTERVENTION FIDELITY

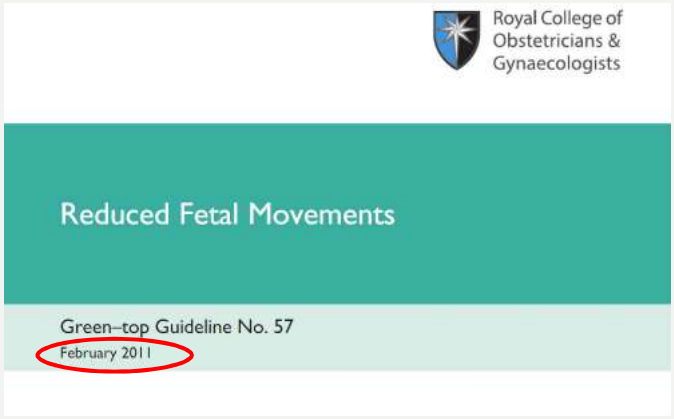
- Intervention fidelity refers to the reliability and validity of the clinical interventions that are used in the randomised trial.
- Fidelity reflects
  - whether the interventions are appropriately performed (application, dosage, and intensity) and whether the interventions adequately represent how the intervention is performed in clinical practice.
- Intervention fidelity is consistently either poorly performed, poorly reported or both.
- There is often limited fidelity in the application of behavioral interventions (Cook and Thigpen 2019)



EQUIPOISE

- There should be "genuine uncertainty in the expert medical community about the preferred treatment" before a randomized trial is allowed to be conducted (DeHoop et al 2015) ie there should exist no decisive evidence that the intervention will be superior to existing treatments or effective at all.

AVAILABLE INFORMATION ABOUT SIGNIFICANCE OF FETAL MOVEMENTS FOR BOTH CLINICIANS AND WOMEN



OTHER PROBLEMS WITH THE STEP-WEDGE CLUSTER RCT

Although all clusters will receive the experimental intervention, it does not always mean that all participating subjects will receive the experimental intervention. (DeHoop et al 2015)

A step wedge cluster RCT often does not meet planned sample size (Eichner et al 2019)

IS THE RCT ALWAYS THE "GOLD STANDARD"?

The special status awarded to RCT is unwarranted and which research method is best depends on what we are trying to discover and on what is already known

In the case of stillbirth much is already known from observational research (case-control, cohort) study. These studies are a source of high level evidence which (particularly if pooled: IPD analysis) can result in strong evidence for practice without the need for RCT.

You cannot know how to use trial results without first understanding how the results from RCTS relate to the knowledge that you already possess about the world, and much of this knowledge is obtained by other methods (Deaton and Cartwright 2018)

It is imperative to understand that RCTs are a form of research design and this design is not appropriate for all forms of research needs. For example, rare outcomes are best studied using case-control designs…An observational case-cohort design will better reflect the population, prevalence and downstream influence of harms (Cook and Thigpen 2019).

CAN A RCT DO "HARM"

The "gold standard" or "truth" view does harm when it undermines the obligation of science to reconcile RCT's results with other evidence in a process of cumulative understanding (Deaton and Cartwright 2018)

- *The RFM care package did not reduce the risk of stillbirths. The benefits of a policy that promotes awareness of RFM remains unproven* (Norman et al 2018)

PROFESSIONAL ORGANISATION RESPONSE

- *Based on the findings demonstrated in the ARRIVE trial, it is reasonable for obstetricians and health-care facilities to offer elective induction of labor to low-risk nulliparous women at 39 weeks gestation. However, consideration for enactment of this elective induction of labor intervention should not only take into account the trial findings, but that this recommendation may be conditional upon the values and preferences of the pregnant woman, the resources available (including personnel), and the setting in which the intervention will be implemented. A collaborative discussion with shared-decision making should take place with the pregnant woman. Additionally, as induction of labor involves coordination between the health care provider and the infrastructure in which induction and delivery will occur, it is critical that personnel and facilities coordinate polices related to the offering of elective induction of labor.*

RCT AND LOE

Original Investigation

March 19, 2019

Levels of Evidence Supporting American College of Cardiology/American Heart Association and European Society of Cardiology Guidelines, 2008-2018

Alexander C. Fanaroff, MD, MHS[1]; Robert M. Califf, MD[2,3,4]; Stephan Windecker, MD[5]; et al

Author Affiliations | Article Information

JAMA. 2019;321(11):1069-1080. doi:10.1001/jama.2019.1122

Editorial Comment

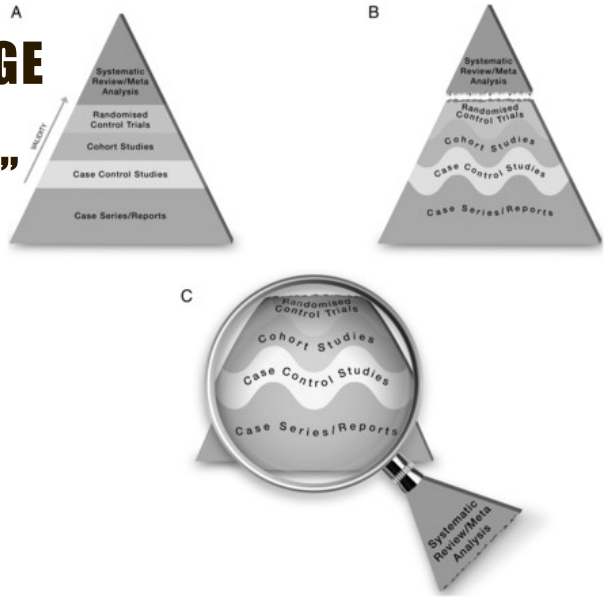
Key Points

Question What proportion of recommendations in current American College of Cardiology/American Heart Association (ACC/AHA) and European Society of Cardiology (ESC) guidelines are supported by evidence from multiple randomized controlled trials (RCTs), and how has this changed over the past 10 years?

Findings In this systematic review of 51 current guideline documents that included 6329 recommendations, 8.5% of recommendations in ACC/AHA guidelines and 14.3% of recommendations in ESC guidelines were classified as level of evidence A (supported by evidence from multiple RCTs), compared with 11.5% of recommendations in a systematic review of ACC/AHA guidelines conducted in 2009.

DO WE NEED TO CHANGE HOW WE VIEW THE EVIDENCE "PYRAMID?"

- The proposed new evidence-based medicine pyramid. (A) The traditional pyramid. (B) Revising the pyramid: (1) lines separating the study designs become wavy (Grading of Recommendations Assessment, Development and Evaluation), (2) systematic reviews are 'chopped off' the pyramid. (C) The revised pyramid: systematic reviews are a lens through which evidence is viewed (applied).



Murad et al 2016

YES!

REFERENCES

- Cook, C.E. and Thigpen, C.A., 2019. Five good reasons to be disappointed with randomized trials. Journal of manual and manipulative therapy
- Deaton, A. and Cartwright, N., 2018. Understanding and misunderstanding randomized controlled trials. Social Science & Medicine, 210, pp.2-21.
- de Hoop, E., van der Tweel, I., van der Graaf, R., Moons, K.G., van Delden, J.J., Reitsma, J.B. and Koffijberg, H., 2015. The need to balance merits and limitations from different disciplines when considering the stepped wedge cluster randomized trial design. BMC medical research methodology, 15(1), p.93.
- Eichner, F.A., Groenwold, R.H., Grobbee, D.E. and Rengerink, K.O., 2018. Systematic review showed that stepped wedge cluster randomized trials often did not reach their planned sample size. Journal of clinical epidemiology.
- Faber J, Fonseca LM. How sample size influences research outcomes. Dental Press J Orthod. 2014 July-Aug;19(4):27-9. DOI: http://dx.doi.org/10.1590/2176-9451.19.4.027-029.ebo
- Flenady, V., Ellwood, D., Bradford, B., Coory, M., Middleton, P., Gardener, G., Radestad, I., Homer, C., Davies-Tuck, M., Forster, D. and Gordon, A., 2019. Beyond the headlines: Fetal movement awareness is an important stillbirth prevention strategy. Women and Birth, 32(1), pp.1-2.
- Grobman, W.A., Rice, M.M., Reddy, U.M., Tita, A.T., Silver, R.M., Mallett, G., Hill, K., Thom, E.A., El-Sayed, Y.Y., Perez-Delboy, A. and Rouse, D.J., 2018. Labor induction versus expectant management in low-risk nulliparous women. *New England Journal of Medicine*, *379*(6), pp.513-523.
- Murad, M.H., Asi, N., Alsawas, M. and Alahdab, F., 2016. New evidence pyramid. BMJ Evidence-Based Medicine, 21(4), pp.125-127.
- Norman, J.E., Heazell, A.E., Rodriguez, A., Weir, C.J., Stock, S.J., Calderwood, C.J., Burley, S.C., Frøen, J.F., Geary, M., Breathnach, F. and Hunter, A., 2018. Awareness of fetal movements and care package to reduce fetal mortality (AFFIRM): a stepped wedge, cluster-randomised trial. The Lancet, 392(10158), pp.1629-1638.
- Tveit, J.V.H., Saastad, E., Stray-Pedersen, B., Børdahl, P.E., Flenady, V., Fretts, R. and Frøen, J.F., 2009. Reduction of late stillbirth with the introduction of fetal movement information and guidelines–a clinical quality improvement. BMC pregnancy and childbirth, 9(1), p.32.