

Is the randomised controlled trial the best?



The randomised controlled trial (RCT) is recognised as the gold standard of research methods, particularly to test efficacy. The primary benefit of the RCT, as everyone knows, is to prevent patient selection bias. And it should also guarantee some rigour of research methodology. It is always prospective.

In a non-randomised study, in a matched-case controlled trial, for example, a researcher may claim to have found for each case another patient who matches for age, weight, smoking habits, parity, or all of the above and more, and so is the perfect comparison to test the outcome. But that, in the case control study, makes one very unreasonable assumption – that the researcher has a perfect understanding of the biology and knows *all* the variables that may affect or ‘confound’ the outcome. That assumption is very often not true.

In a case control study, a researcher who is biased in ‘control’ selection may be selective in case selection or exclusion also. Case benefit is exaggerated. How do we know that these influences in trials are important? Because intervention benefit is shown in 60% of case control studies and only 25% of RCTs – a good reason for the RCT to be the standard bearer.

But it is not that simple. There are many reasons to criticise an RCT. RCT errors relate to which patients are put into an RCT, from which universal applicability is to be presumed, or to who is taken out of the analysis, or to the exaggeration of effect; in a large study, for example, negligible or small difference *will* achieve statistical significance. For example, the treatment of minimal endometriosis by ablation or excision allegedly significantly improves fertility. RCTs show this. If RCTs are small, their numbers can be accumulated in a meta-analysis of randomised trials so that the Peto plot comparing likelihood of benefit is clearer. This collection of RCTs allegedly discards a single RCT, the Rolls Royce, for a fleet of Rolls Royces. And yet, look at the meta-analysis data for endometriosis carefully and some 80% of patients will not obviously benefit.

The idea that an accumulation of small RCTs can be meta-analysed or synthesised into a pure weight of evidence, all with their individual procedural problems and often comparing slightly different outcomes, is virtual alchemy, and highly contested. And yet this is ‘Grade 1A’ evidence.

Another major weakness of RCT meta-analyses that do demonstrate an effect is that those particular RCTs included in the process that show no benefit are acknowledged and then ignored.

Randomisation in RCTs can be tampered with; patients may not simply be allocated into treatment or non-treatment by opening a sealed letter or allocation of a randomly generated computer number as they *randomly* walk in, but may be allocated into groups not previously specified, and some may be excluded. This can be tested by the claimed incidence of the condition at that unit – the numbers are too small. It takes deep investigation to spot the simple deceit. Who deceives? Zealous practitioners/researchers who believe they know what the outcome *should* be.

Also, if an unusual population is entered into an RCT to then illustrate an effect in a wider population, extrapolation may not be possible. In the Women’s Health Initiative study, the mean age at

randomisation was 64 years; yet most often hormone replacement therapy (HRT) is started perimenopausally. The results may not be relevant to a standard HRT population, for very good reasons.

The geographical variation of selected patients is also important. A liver RCT on the hepatitis B virus in Tokyo is likely not to have validity in Johannesburg; in Tokyo the virus causes hepatocellular carcinoma, in Johannesburg it does not (but then again, HIV changes everything). Similarly, resection of gastric carcinoma provides a clear survival benefit in the Far East not repeated elsewhere.

How do patients get inappropriately excluded from RCTs? One method is to clean the data by excluding all subjects who do not *correctly* complete a protocol (‘per protocol’ analysis) rather than all those who start the trial (‘intention-to-treat’ analysis). If a manufacturer wishes to exaggerate the benefit of the application of the HPV vaccine, it will use the former, not the latter. A manufacturer may even compare its pure ‘per protocol’ data to uncleaned data of a rival. This is clearly ‘unreal’.

Also, the presumed relevance of a variable tested in an RCT, the actual fundamental of the RCT, may be incorrect (we already know that we have little understanding of variables from the case control v. RCT comparison of demonstrated effects). One such controversy of the true variable surrounds HRT; this controversy is centred on lipids. ‘HRT benefits cardiovascular risk,’ stated the highly credible case control studies of the 1970s, 80s and 90s. These studies have famous names – The Walnut Creek Study, The Nurses’ Health Study. They had thousands of patients and were/are very expensive.

Lipids are changed by HRT – low-density lipids down, high-density lipids up, etc. – a favourable trade. Lipids are connected to atherosclerosis, atherosclerosis to infarction. Manufacturers wishing to introduce new drugs ran small RCTs comparing lipid effects of new drugs with those of old, established ones. If the results were favourable, the new drugs should be purchased; and the RCT, the champion testing vehicle, said so.

But here is a major flaw – nobody knows *how* HRT may benefit cardiovascular risk in that younger-initiated patient. It may not be lipids at all. If it were, that effect would be sustained (through retarded vessel plaque formation) after cessation of treatment; in many studies it simply is not. So the chosen RCT variable could be wrong, thereby invalidating the RCT. Debate rages over the possible significance of oestrogen as a vasodilator (strange, since its absence causes flushing). That *is* reversible on cessation of treatment, and not tested in any RCTs observing a potentially incorrect though plausible lipid fluctuation.

In obstetrics, a highly unexpected possible true variable was the suggested physiological benefit of the single attending midwife in labour, and not the aggressive syntocinon augmentation, in the highly successful and not repeated technique of the Dublin Rotunda of the ‘active management of labour’. Many RCTs tested aggressive v. non-aggressive augmentation. If the theory stands – wrong variable, and so invalid RCTs to challenge the Rotunda technique (the Rotunda also ascribed the benefit to syntocinon). Curiously, the attribution of this theorised ‘attendant’ variable may have significantly altered the acceptance of patient support in labour.

There are three famous authors in the obstetrics and gynaecology literature who have opposed particular RCTs for very different reasons to those stated above. They are Liggins and Howie, the researchers on steroids to enhance fetal lung maturity (and other effects), who were reluctant to be involved in a multicentre RCT of antenatal steroids, stating that the overwhelming evidence made any placebo arm unethical; and Smithells, the UK proponent of folate supplements to prevent neural tube defects, who said the same about the multicentre Medical Research Council vitamin supplement RCT.

Whatever the limitations and pitfalls of RCTs, their recruitment, their fundamentals, their analysis and interpretation, certainly one aspect can be championed for certain – most often they are the primary prover, or disprover, of effect. But they must be adequately

large, correctly and properly explained, and correctly interpreted – fastidiously and with limitations.

However great the ranking of the RCT in efficacy testing, in the revelation of complications and side-effects it is the larger cohort-based studies and post-marketing surveillance that remain predominant. Tragically, these data are not always widely distributed, but thankfully, often they are. And this is one example of data ‘missed’ by many a randomised trial.

William Edridge

Editor

william.edridge@gmail.com

S Afr J OG 2014;20(3):74-75. DOI:10.7196/SAJOG.943