Deconstructing clinical trials: A guide for junior doctors

Jun 21, 2020 | 0 💬 | General medicine, ontheblogs, research

Author Aran Kanagaratnam

Editors Abhijit Pal | Antonia Clarke

Clinical trials – Randomised controlled trials

The randomised controlled trial (RCT) is one of the most important medical advances of the last century. RCTs are the 'gold standard' for comparing a new diagnostic test or treatment with the existing standard of care – and for good reason. RCTs are carefully controlled experiments to evaluate whether a new treatment is superior to the existing standard of care. Researchers define a trial population through inclusion and exclusion criteria, set a hypothesis, prespecify primary endpoints that need to be met to disprove the null hypothesis, and a statistician will help determine the sample size required in order to test the hypothesis based on the estimated effect size.

RCTs are carefully designed in order to avoid bias – for example, randomisation will seek to eliminate selection bias (read more - https://catalogofbias.org/biases/selection-bias/) and single or double blinding will seek to reduce performance bias (read more https://catalogofbias.org/biases/performance-bias/).

But not all RCTs have equal value to clinical practice. Some have deep methodological flaws that mean their results cannot be transferred to answer clinical questions. From choosing inappropriate patient populations to using unsuitable endpoints, and all the biases in between – it is vital for doctors to learn skills in critically analysing clinical trial design and reporting. This can be utterly confusing for doctors in training, especially as much of our training is learning accepted practice before beginning to question the status quo.

The importance of research for junior doctors

Understanding (and conducting) research is more important for junior doctors than ever before. A recent review of Australian specialist colleges found that while almost all required trainees to carry out a research project, only 1 in 5 required formal research training (1). The gap between the expectation of being adept at critically appraising research and the lack of formal training is a major reason why so many junior doctors find research daunting. Moreover, at the senior level, doctors need critical appraisal skills so they can determine whether new research is relevant to and should change their clinical practice.

This guide aims to provide a simple framework for critically analysing clinical trials, as well as some common biases and pitfalls that feature in many trials. We'll talk about the four questions you should ask yourself whenever you read a clinical trial.

1. What question was the study trying to answer? (and is that the right question?)

Let's suppose you are asked to see Mrs Jones, a 65-year-old lady who is recovering from a large MCA stroke and is ready to be discharged home. You remember hearing your boss talk about a very recent trial of NewDrug, a new medication that is supposed to prevent strokes. How can you work out whether Mrs Jones should take NewDrug?

The first step is working out what the point of the study was. A useful framework for breaking down RCTs is PICO – who was the **population** studied, what was the **intervention** and **comparator** and what **outcome** measure did the trial use? Often just looking at the title can provide a lot of information: a (fictional) study entitled 'A randomised controlled trial of NewDrug vs CurrentDrug in secondary prevention of strokes in high-risk patients' tells you what medication the intervention and control arms got and that it was for secondary prevention of strokes.

Some of the specifics, however, are still unanswered, and need to be found within the text. What dosing regimes were used for each drug? Who exactly are 'high-risk' patients? How long was the follow-up? We want answers to these questions because the answers highlight how relevant the study is to our patient, Mrs Jones. Good RCTs should explicitly document who they recruited and how, exactly what the intervention and control arms were and what the outcomes were. A comprehensive checklist of what should be reported for an RCT can be found in the CONSORT statement guidelines (2).

Once we know what question the study was trying to answer, we need to see if that matches the question we want answered. Roughly speaking, this question is some variation of 'for a patient group that I see in my practice, is this new intervention better than my current care, as measured by a meaningful endpoint?' We can split this up into asking whether the appropriate patient population, intervention and endpoints have been chosen.

2. Is this the trial patient population relevant to my clinical practice?

Selecting the right patient population for eligibility in a clinical trial is complex. Too restrictive and the study population does not reflect the real world and will be clinically irrelevant. Yet too inclusive and then we have combined together groups that have different responses to the intervention, and a true effect may be missed. The study result then reports an average effect size that does not tell you which groups actually benefit from it.

Suppose our fictional trial had an inclusion criterion that participants had to be able to run ten kilometres. We would then have a problem of generalisability – where the results of this trial might show that NewDrug works in the very fittest post-stroke patients, but not necessarily in anyone else. This problem is common in clinical trials, where strict inclusion criteria result in a patient population much healthier than the real-world cohort with the same disease. This a key lesson – the effect size observed in a clinical trial is often larger than that observed in real-world clinical practice as the patients, on average, are less well than patients in a clinical trial.

On the other hand, selecting a patient population that is too heterogeneous has its own problems. If in our trial half the patients are very high risk and significantly benefit from NewDrug, but the other half are very low risk and do not benefit, we may end up with a trial result that combines them and says overall everyone moderately benefits from the intervention, which is not, in fact, true for anyone. Sometimes analysing by subgroup can highlight different response rates to the intervention, but for this to work we need to know which subgroups are likely to be different from each other. Additionally, trials are statistically powered only for their primary endpoint, and thus subgroup analyses are exploratory only.

Avoiding selection bias during clinical trials

The treatment group and control groups can be poorly balanced. Selection bias is when doctors recruit selectively based on what they perceive the next treatment allocation to be, such as waiting to enrol a sick patient to ensure they get the intervention. This issue can be addressed with double blinding, so neither researchers nor participants know which treatment is being allocated, and by having a proper randomisation procedure, which should be detailed in the methods section. The intervention and control groups should be similar at baseline, as would be expected with proper randomisation, and any differences should be accounted for. One way this can be achieved is with stratified randomisation, where participants are divided based on a homogeneous patient characteristic that is expected to have an impact on the effect size e.g. age (creating a 'strata') and then an appropriate number of participants are randomised from each strata.

3. Were the intervention and control appropriate?

One of the easiest ways the efficacy of a new intervention can be artificially inflated is by making the control arm worse. Suppose that in our fictional trial the CurrentDrug dosing regime only gave half the dose recommended by current clinical guidelines. This would then artificially boost the relative efficacy of the NewDrug, but would not tell us whether the NewDrug was actually better than our current standard of care. It is crucial that triallists give the standard of care therapy to the control patients so that the clinical community can truly work out if the new therapy is superior. This is not simple – if it takes 5 years for an RCT to complete recruitment and follow up, the standard of care may have changed which makes the results harder to interpret.

Unequal loss to follow-up can bias results. If NewDrug required administration under physician supervision and only half the recruited participants adhered, these patients may be systematically different from non-adherers, e.g. healthier, and this so-called attrition bias may skew the results of the study. Most RCTs thus use an "intention-to-treat" analysis (https://litfl.com/intention-to-treat-analysis/), where all randomised participants are included in follow-up, even those that fail to adhere. Follow-up should be the same for both groups, and the only difference being the intervention itself. Tolerability and adverse events associated with the intervention should also be noted.

4. Are the endpoints meaningful?

Ideally, the endpoints of RCTs should represent clinical events that we care about. In our fictional trial, a clinically meaningful endpoint might be strokes within 2 years. Due to resource constraints, surrogate outcomes may be used. These are outcomes that ideally correlate with clinical outcomes and require shorter follow-up to maintain statistical power. For example, in a cancer trial of a new therapy, tumour shrinkage could be measured instead of overall survival. The potential benefit is that trials could maintain statistical power while being shorter and cheaper, but the risk is that the surrogate endpoint is a poor proxy for the clinically relevant endpoint (overall survival).

Sometimes endpoints are a composite of clinical outcomes. This can lead to a situation where one measurable component of the composite comprises most of the event rate. In a recent trial comparing percutaneous coronary intervention with medical therapy in patients with stable coronary artery disease, the endpoint was a composite of urgent revascularisation, myocardial infarction and death (3). There was a significant difference between groups (27.0% vs 13.9%) but this was driven entirely by the difference in urgent revascularisation (21.1% vs 6.3%), and there was no difference between the groups for either myocardial infarction or death. Here, despite a positive trial result, the difference was driven by a surrogate endpoint, with no difference in the more clinically meaningful clinical endpoints. While this in itself does not automatically negate the trial's relevance, it should make you consider whether the surrogate endpoint is meaningful.

Statistical vs clinical effect of clinical trials

Finally, the treatment effect size needs to be contextualised. A result that is statistically significant does not necessarily imply it will be clinically significant. If NewDrug prevents one extra stroke for every thousand people treated for five years it is unlikely that it is worth changing current practice for, even if the result is statistically significant. This is particularly important when considering costly or complex treatments. One example is a trial in pancreatic cancer of erlotinib and gemcitabine vs gemcitabine alone (4). The intervention added ten days to a patient's survival – a statistically significant difference, but likely clinically irrelevant. On top of this, the regime would have cost an additional US$15000 (5).

Appraising clinical trials takes practice

As with the rest of medicine, the key to appraising clinical trials is practice. This framework is just a starting point. Discussing new studies with senior colleagues can provide a different perspective on whether it is worth paying attention to certain trials. At the very least, understanding this framework should reduce the chance of you being completely lost at next month's journal club.

−

References

1. **Stehlik P, Noble C, Brandenburg C, Fawzy P, Narouz I, Henry D, Glasziou P. How do trainee doctors learn about research? Content analysis of Australian specialist**

colleges' intended research curricula. BMJ open. 2020 Mar 1;10(3):e034962.

2. Schulz KF, Altman DG, Moher D, Consort Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. Trials. 2010 Dec 1;11(1):32.

3. Xaplanteris P, Fournier S, Pijls NH, Fearon WF, Barbato E, Tonino PA, Engstrøm T, Kääb S, Dambrink JH, Rioufol G, Toth GG. Five-year outcomes with PCI guided by fractional flow reserve. New England Journal of Medicine. 2018 Jul 19;379(3):250-9.

4. Moore MJ, Goldstein D, Hamm J, Figer A, Hecht JR, Gallinger S, Au HJ, Murawa P, Walde D, Wolff RA, Campos D. Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advanced pancreatic cancer: a phase III trial of the National Cancer Institute of Canada Clinical Trials Group. Journal of clinical oncology. 2007 May 20;25(15):1960-6.

5. Miksad RA, Schnipper L, Goldstein M. Does a statistically significant survival benefit of erlotinib plus gemcitabine for advanced pancreatic cancer translate into clinical significance and value?. Journal of Clinical Oncology. 2007 Oct 1;25(28):4506-7.

Related Podcasts

- Why Junior Medical Officers should get involved in research

Related Blogs

- Getting involved in research

- How to document well

Tags: #clinical trials,#CONSORT statement,#critical appraisal,#diagnostic test,#medical research,#PICO,#randomised controlled trial,#RCT,#research,#tipsfornewdocs,#treatment