

Summary of the reasons why Randomised Controlled Trials(RCTs) provide the most trustworthy sort of evidence about whether a treatment works.

By Amanda Burls

Why use a trial? A trial means an experiment. It is important to show experimentally that an intervention works. Just because something seems like a good idea it does not necessarily mean that it is a good idea. It is important to actually carry out the intervention and see whether it helps people make sense of the evidence or not.

Why should it be controlled? Controlled means there is a comparison group. It is important to have a comparison group which does not get the intervention because people could have achieved the required outcome without having anything to do with the treatment. The study should be able to demonstrate if there is improvement **over and above** any background improvement there might be. (Note "controlled" in this context does not mean that it was well organised and standardised although a good trial should also have these characteristics.)

Why should it be placebo-controlled? The comparison group should receive some kind of dummy treatment rather than nothing. This is so we can be more sure that any differences between the intervention and control groups were **because** of the intervention and were not due simply to the non-specific effect of giving people attention or their optimism that the intervention would work (placebo effect).

Why should it be randomised? This means that whether an individual receives the active intervention or the control intervention is decided completely by chance. The reason for choosing to randomly allocate people to either group is because **it is important that people in both the intervention and control groups are similar** at the start of a trial. If they are not similar, the differences at the end of the trial might simply be due to differences that existed at the start of the trial (selection bias) and have nothing to do with the intervention given. For example, if we compared those who want a treatment with those who do not, any improvement might simply be due to the enthusiasm of the group volunteering and not to the intervention (volunteer bias is one sort of selection bias). Experience shows this is the best way to get groups that are similar. It is important that the doctor or researcher should not know or be able to guess which group a patient will be randomised to as this can inadvertently introduce selection bias. Preventing the researcher knowing is known as "allocation concealment".

Why should it be blinded? The person assessing the outcome of the trial should be blinded to whether the participants had been in the intervention or control group. This is to prevent the assessor consciously or sub-consciously assessing people differently (measurement bias). Where possible, the participants in a trial should themselves be blinded as to whether they are receiving the active or control intervention in order to reduce bias due to the placebo effect. Allocation concealment also helps ensure that the researchers are blinded as to which patients are receiving active treatment.

What is intention to treat analysis? When the results are analysed people should be analysed in the group they were originally randomly allocated to. So that if, for example, a person allocated to receive the control intervention actually received the intervention, they would nonetheless be analysed as if they had not had the intervention. Similarly, if someone allocated to the control intervention did not do so, they would still be counted with the intervention group. If results are not analysed in this way, the whole purpose of the original randomisation breaks down: the people in the two groups may then systematically differ from each other and any differences in the results of the two groups could simply be due to this fact. Using intention to treat analysis only weakens the observed effect of an intervention but does not make an intervention appear

effective when it was not. That is to say, it does not undermine any observed association between an intervention and outcome - it makes it **more** believable.

Can I trust this research result?

Lots of studies claim to demonstrate that a particular intervention “produced” a particular result; that a treatment was effective. Caution is needed before you can draw this conclusion from a study:

When a study demonstrates an association between an intervention and particular outcome (e.g. it may show that people are twice as likely to survive a heart attack if they are given a particular drug just after the heart attack happens), this **might be** because the drug is effective **but the result can be due to a number of different reasons**. It is important to look closely at the research to exclude these reasons before concluding that it was the treatment that probably produced the result.

Other reasons for apparent associations include:

- bias
- chance
- other explanations (confounding)

Why do we look at bias?

Bias can creep into a study in several ways but a well-designed and well-conducted study helps minimise these biases. If a study is seriously biased then **there is no point going any further** and looking at the results because you cannot trust the results. So that the first thing to do when looking at any study is to check that the methods that were used are going to produce trustworthy results. If the methods are weak then the study will be biased and untrustworthy (it will not be valid). CASP workshops give a structured way of looking at studies to help you detect ones that are biased.

Could the result have happened by chance?

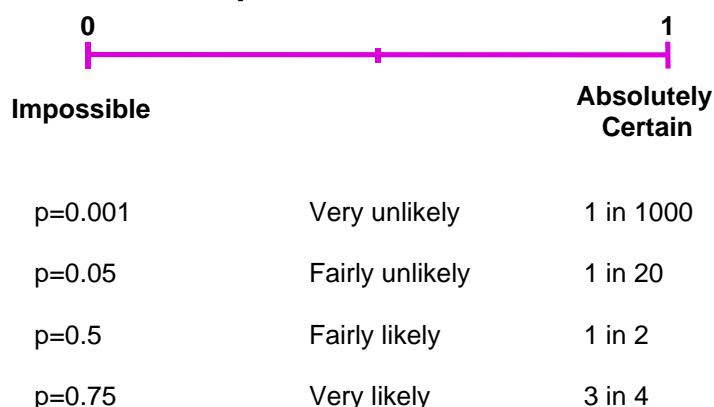
Once you have a study with a sound methodology that minimises bias (like an RCT), it is still possible that any association demonstrated might simply have happened by chance (and that if you repeated the study you would not find the same association).

We can apply statistical tests to help us decide whether the result could simply be due to chance. When people say that a study is “statistically significant” they are simply saying that it is unlikely that such a result would be seen by chance. (A study can be statistically significant without being clinically significant or important.)

How unlikely to have occurred by chance does a study have to be to be considered statistically significant? There is no one answer to this question - it is a matter of judgement. Many researchers use a probability (p-value) of less than 0.05 as the cut off for “statistical significance”, i.e. when the sort of result seen in a study would occur by chance less than once in 20 studies.

If you remember that probability (p-value) can only take values between 0 and 1 then this can help you interpret how likely or unlikely something is.

The p-value in a nutshell

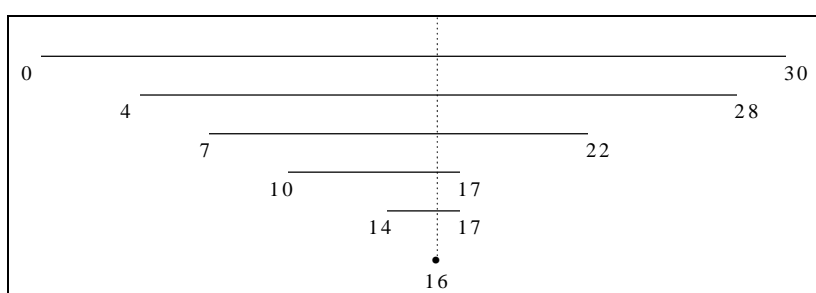


Confidence Intervals (CI)

The confidence interval gives the range in which you think the real answer lies with a given degree of certainty.

If you have a bag (which you cannot see into) containing thirty sweets that could be either orange or white, then the number of orange sweets in the bag at the start could lie between nought and thirty, i.e. the 100% confidence interval for the number of orange sweets is 0 - 30 (100%CI = 0 - 30).

If you draw a handful of sweets out of the bag and there are 4 orange and 2 white sweets in your hand then you know that there were at least four orange sweets at the start and no more than 28 (there were 30 sweets altogether and you now know that at least 2 were white), i.e. your confidence interval for the number of orange sweets there were at the start is now 4-28. If you grab another handful of sweets (without replacing the first handful) and pull out 3 orange and 6 white sweets then the confidence interval for the number of orange sweets in the bag at the start is 7-22. The next selection contains 3 orange and 5 white sweets, giving you a CI of 10-17 for the number of orange sweets at the start. The next selection contains just 4 orange sweets giving a CI of 14-17. The last three sweets are drawn out and two are orange and one is white. You conclude that there were 16 orange sweets at the start.



Notice that, as the total sample of sweets drawn out gets larger, so the CIs get shorter. In this example we calculated absolute confidence intervals, i.e. at every point we were absolutely sure that the real number of orange sweets lay somewhere between the two confidence limits.

In real life, when thinking for example about interventions to improve health, things get a little more complicated. In the sweet example, there was a fixed, finite, determined, number of sweets in the bag. However, when we think about interventions there is no predetermined (finite) number of things we are talking about. For example, we may be interested in giving steroids to women in pre-term labour to help their babies breathe more easily when they are born, but this includes

women in the future who may **not yet** be pregnant or have gone into labour. In most of the circumstances we are interested in we can never be **absolutely** sure of the truth because we only ever see a sample of the possible cases. This doesn't matter: we do not need to be absolutely sure, we just need to be **reasonably** sure. Most studies use 95% confidence intervals, i.e. they give the range where we expect that the true result will lie 95% of the time (i.e. only in 1 in 20 studies on average will the real value lie outside the confidence limits).

On a blobbogram if the 95% CI does not cross 1 (the line of no treatment effect) then the result is statistically significant at $p=0.05$. ("Blobbogram" is CASP's name for an odds ratio or Forest plot.)