

# Critically appraising randomised trials

**Derek Richards**

*Centre for Evidence-based Dentistry, Oxford.*

**Critical appraisal is one of the key skills of evidence-based practice and is now increasingly being taught in dental schools. Here we outline the key principles of appraising randomised controlled trials (RCT).**

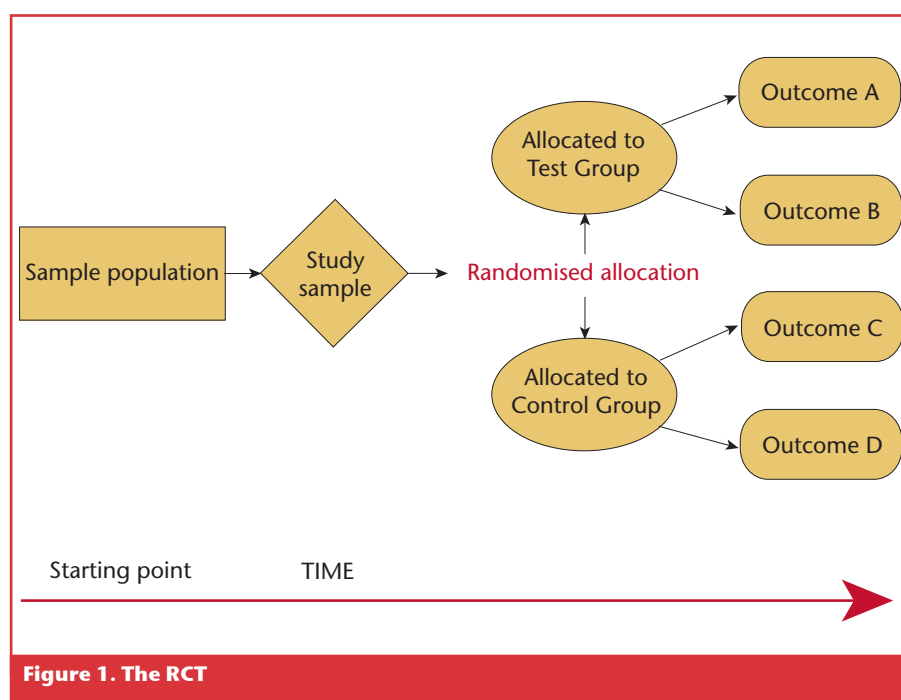
When testing a new treatment in a clinical trial, there are three possible explanations for why it did or did not work as expected - chance, bias or the 'truth'. Bias and chance (or random error) can be reduced through appropriate study design. The RCT is a simple but powerful research tool and it is the best way of assessing whether a cause and effect relationship exists between the treatment and outcome. Although the first controlled trials are thought to have taken place over 300 years ago ([www.jameslindlibrary.org/context/principles-of-testing.html](http://www.jameslindlibrary.org/context/principles-of-testing.html)), their use began in earnest with the pioneering work of Bradford Hill in the late 1940s.<sup>1</sup>

The key feature of the RCT, randomisation (Figure 1), if done properly, keeps the study groups as similar as possible, thus enabling investigators to separate and measure the effect of the intervention and reduce selection bias. Random allocation does not protect RCT against other types of bias. Furthermore, although RCT are powerful tools, their use is limited by ethical and practical concerns, and they tend to be comparatively costly. Because of these constraints, there is the risk that design compromises or poor reporting of RCT will impact on these studies' quality. It is important, therefore, that those who read published reports of RCT are able to appraise their quality.

Even though an RCT is an appropriate tool to evaluate a healthcare intervention, readers should be aware that of the many important health issues that could be studied by using an RCT, as yet, many have not been. In addition, even when RCT are available, the data they provide may be insufficient to provide all the answers required by clinicians, patients or policymakers.

### Bias in RCT

Bias has been defined as a systematic error, or deviation from the truth, in results or



**Figure 1. The RCT**

**Table 1. Types of bias in RCT**

Type of bias	Description
Selection bias	Systematic differences between baseline characteristics of the groups that are compared
Performance bias	Systematic differences between groups in the care that is provided, or in exposure to factors other than the interventions of interest
Attrition bias	Systematic differences between groups in withdrawals from a study
Detection bias	Systematic differences between groups in how outcomes are determined
Reporting bias	Systematic differences between reported and unreported findings

inferences. Biases can operate in either direction leading to underestimation or overestimation of the true intervention effect. The Cochrane Collaboration Handbook ([www.cochrane-handbook.org](http://www.cochrane-handbook.org)) classifies biases into selection bias, performance bias, attrition bias, detection bias and reporting bias (Table 1).

**Selection bias** As noted above, if randomisation is done correctly, selection bias can be prevented. Successful randomisation

requires a clearly specified chance (random) process known as sequence generation, and strict implementation of random allocation or allocation concealment. For example, this could be a computer-generated random sequence concealed from those involved in enrolment into the trial.

**Performance bias** Effective blinding (or masking) of both the study participants and personnel should ensure that the groups being compared will receive similar amounts

**Table 2. Main appraisal questions from the Centre for Evidence-based Medicine (CEBM) and Critical Appraisal Skills Programme (CASP) worksheets**

CASP appraisal questions	CEBM appraisal questions	Purpose of question
Did the study ask a clearly focused question?	What question did the study ask?	Validity question
Was this a RCT and was it appropriately so?	Was the assignment of patients to treatments <u>randomised</u> ?	Validity question
Were participants appropriately allocated to intervention and control groups?	Were the groups <u>similar</u> at the start of the trial?	Validity question
Were participants, staff and study personnel 'masked' to participants' study group?	Aside from the allocated treatment, were groups treated equally?	Validity question
Were all of the participants who entered the trial accounted for at its conclusion?	Were all patients who entered the trial accounted for? – and were they analysed in the groups to which they were randomised?	Validity question
Were the participants in all groups followed up and data collected in the same way?	Were measures <u>objective</u> or were the patients and clinicians kept ' <u>blind</u> ' to which treatment was being received?	Validity question
Did the study have enough participants to minimise the play of chance?		Validity question
How are the results presented, and what is the main result?	How large was the treatment effect?	Results question
How precise are these results?	How precise was the estimate of the treatment effect?	Results question
Were all important outcomes considered so that the results can be applied?	Will the results help me in caring for my patient?	Relevance question
See <a href="http://www.phru.nhs.uk/Pages/PHD/resources.htm">www.phru.nhs.uk/Pages/PHD/resources.htm</a> for full appraisal worksheet	See <a href="http://www.cebm.net/index.aspx?o=1157">www.cebm.net/index.aspx?o=1157</a> For full appraisal worksheet	

RCT, Randomised controlled trial.

of attention, additional treatment and tests. Blinding also protects the randomisation sequence after allocation. Blinding is not always possible, however, such as in the case of surgery.

**Detection bias** Blinding is also important for those who measure the outcomes, particularly for subjective outcomes such as the gingival index.

**Attrition bias** Missing outcome data may be caused by participant drop-out (attrition) eg, individuals missing a final appointment or moving away from an area, or they may have been excluded from the analysis in some way. Intention-to-treat analysis is often used to deal with missing data and is considered to be the least biased method of assessing intervention effects, but it does not have a clear and consistent definition.<sup>3</sup>

**Reporting bias** Reporting bias could result from selective omission of outcomes, selected choice of data, selective reporting of analyses or subsets of data and under-reporting of data.

The biases noted above are the most important ones affecting RCT, although there are a number of other biases that could impact on study quality. Some study designs, eg, crossover trials, cluster-RCT, and trials with multiple intervention groups, have other potential biases, as do those with

baseline imbalances or instances when a trial stops early.

**Appraisal questions and checklists**

A number of critical appraisal questions or checklists are available (see Table 2).

There are a number of textbooks and websites that provide information on critical appraisal (for links, see [www.cebm.org/practising-ebd/appraise/resources-for-appraising/](http://www.cebm.org/practising-ebd/appraise/resources-for-appraising/)). Despite the range of materials available for appraisal of papers there are only three essential questions that need to be asked of any paper:

- Is the study valid?
- What are the results?
- Are the results relevant?

**Is the study valid?**

What we are trying to ascertain here is whether the study was conducted properly. From Table 2, we can see that both the Centre for Evidence-based Medicine (CEBM) and Critical Appraisal Skills Programme (CASP) have a number of questions related to studies' validity. The first of these addresses the issue of whether the researchers are clear about the question they are asking. The subsequent validity questions are designed to address the key biases affecting RCT, as noted above. Both the CASP and CEBM worksheets provide additional information to help the user

answer these questions. The CASP worksheet provides hints of what to look for whereas the CEBM worksheet identifies what is best and where that information is most likely to be found in the paper.

**What are the results?**

Results can be described in a variety of way depending on the outcomes considered, but are often presented as dichotomous outcomes, eg, caries free or restoration failure. We might consider a study of a caries preventive agent in which 15% ( 0.15) of the test group patients remained caries-free, but only 10% (0.10) of the control group at 2 years. The results could be presented in several ways.

Relative risk

$$= \frac{\text{risk of the outcome in the treatment group}}{\text{risk of the outcome in the control group}}$$

$$= \frac{0.15}{0.10}$$

$$= 1.5$$

The relative risk (RR) tells us how many times more likely it is that an event will occur in the treatment group relative to the control group. An RR of 1 means that there is no difference between the two groups, ie, the treatment had no effect. An RR <1 means that the treatment decreases

## TOOLBOX

the risk of the outcome. An RR >1 means that the treatment increased the risk of the outcome.

In our example the RR = 1.5 (>1) so the treatment increased the chances of being caries-free.

### Absolute risk reduction

Absolute risk reduction (ARR) = risk of the outcome in the control group – risk of the outcome in the treatment group.

$$= 0.10 - 0.15 = -0.05 \text{ or } -5\%$$

(The figure is negative because a good outcome here is a reduction in caries.)

The ARR tells us the absolute difference in the rates of events between the two groups and gives an indication of the baseline risk and treatment effect. An ARR of 0 means that there is no difference between the two groups; thus, the treatment had no effect. The absolute benefit of treatment is a 5% improvement in the number of caries-free patients.

### Relative risk reduction

Relative risk reduction (RRR)

$$= \frac{\text{ARR}}{\text{risk of the outcome in the control group}}$$

An alternative way to calculate the RRR is to subtract the RR from 1 (eg, RRR = 1 – RR).

In our example, the

$$\begin{aligned} \text{RRR} &= 0.05/0.1 &= 0.5 \text{ or } 50\% \\ \text{or RRR} &= 1 - 1.5 &= 0.5 \text{ or } 50\% \end{aligned}$$

The RRR is the complement of the RR and is probably the most commonly reported measure of treatment effects. It tells us the reduction in the rate of the outcome in the treatment group relative to that in the control group. In our example, treatment reduced the risk of caries development by 50% relative to that in the control group.

### Number needed to treat

Number needed to treat (NNT) is the inverse of absolute risk reduction and is calculated as 1/ARR.

In our example, NNT = 1/0.05 = 20

The number needed to treat represents the number of patients we need to treat with the therapy under investigation in order to prevent one bad outcome. It incorporates the duration of treatment. Clinical significance can be determined to some extent by looking at the NNT, but also by weighing the NNT against any harms or adverse effects of therapy.

For the example above, this means that we would need to treat 20 people for 2 years in order to have one extra patient free of caries. Because our study uses a small sample of any given population, the true chance (risk) of the outcome we are interested in occurring in the wider population is not known, so the best that we can do is estimate this, based on our study sample. This is known as a point estimate and we gauge how close this estimate is to the true

value by looking at the confidence intervals (CI) for each estimate. If the CI is fairly narrow then we can be confident that our point estimate is a precise reflection of the population value. The confidence interval also provides us with information about the statistical significance of the result. If the value corresponding to no effect falls outside the 95% CI then the result is statistically significant at the *P* 0.05 level. If the CI includes the value corresponding to no effect then the results are not statistically significant.

### Are the results relevant?

Once you are happy with the trial's validity, you need to decide whether the results can be applied to your patient/s. Key considerations are whether your patients are so different from those in the study that the results cannot apply, if the treatment proposed is feasible in your practice setting, and whether the potential benefits of the treatment outweigh the potential harms of treatment for your patient.

Many people find critical appraisal daunting, but by using appraisal worksheets and regular practice, preferably with a group of like-minded colleagues, it is a skill that can be developed rapidly and one that is core to the evidence-based approach.

*Evidence-Based Dentistry* (2009) **10**, 88-90.  
doi:10.1038/sj.ebd.640673