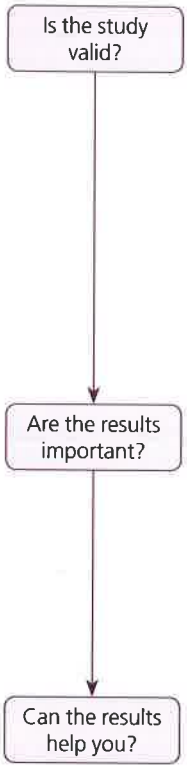
Appraising therapy articles



1 Was there a clearly defined research question?
2 Was the assignment of patients to treatments randomized and was the randomization list concealed?
3 Were all patients accounted for at its conclusion? Was there an 'intention-to-treat' analysis?
4 Were research participants 'blind'?
5 Were the groups treated equally throughout?
6 Did randomization produce comparable groups at the start of the trial?

What is the benefit of the treatment?

RR = EER/CER
RRR = (EER – CER) / CER
ARR = EER – CER
NNT = 1 / ARR

What are the confidence intervals?

How much of the benefit would you expect to see for your patient?

$$\textbf{NNT} = \frac{1}{((\text{PEER} \times \text{RR}) - \text{PEER})}$$

See p. 74.

Is the study valid?

1 Was there a clearly defined research question?

What question has the research been designed to answer? Was the question focused in terms of the population group studied, the intervention received and the outcomes considered?

2 Were the groups randomized?

The major reason for randomization is to create two (or more) comparison groups which are similar at the start of the trial. To reduce bias as much as possible, the decision as to which treatment a patient receives should be determined by random allocation.

Why is this important?

Randomization is important because it spreads all confounding variables evenly amongst the study groups, even the ones we don't know about.

Jargon

Stratified randomization
Often, there are important clinical features which we already know can affect outcomes. If these are not evenly spread amongst the subjects we could end up with a biased result. Patients can be randomized within these categories to ensure that the that these factors are equally distributed in the control and experimental groups.

Block randomization
Block randomization is a technique for ensuring that each of the treatment groups has the right number of participants while retaining allocation concealment.

Allocation concealment
As a supplementary point, clinicians who are entering patients into a trial may consciously or unconsciously distort the balance between groups if they know the treatments given to previous patients. For this reason, it is preferable that the randomization list be concealed from the clinicians.

This is known as allocation concealment and is the most important thing to look for in appraising RCTs (Schulz 1995).

3 Were all patients accounted for at its conclusion?

There are three major aspects to assessing the follow up of trials:

- Did so many patients drop out of the trial that its results are in doubt?
- Was the study long enough to allow outcomes to become manifest?
- Were patients analysed in the groups to which they were originally assigned?

Intention-to-treat

This means that the patients should *all* be analysed in the groups to which they were originally assigned, even if they switched treatments during the trial.

This is important because it's the only way we can be sure that the original randomization is retained, and therefore that the two groups are comparable.

Drop-out rates

Undertaking a clinical trial is usually time-consuming and difficult to complete properly. If less than 80% of patients are adequately followed up then the results should be ignored.

You look at the follow-up rate reported in the study and ask yourself 'what if everyone who dropped out had a bad outcome?'

Length of study

Studies must allow enough time for outcomes to become manifest. You should use your clinical judgment to decide whether this was true for the study you are appraising, and whether the length of follow up was appropriate to the outcomes you are interested in.

4 Were the research participants 'blinded'?

Ideally, patients and clinicians should not know whether they are receiving the treatment. The assessors may unconsciously bias their assessment of outcomes if they are aware of the treatment. This is known as observer bias.

So, the ideal trial would blind patients, carers, assessors and analysts alike. The terms 'single-', 'double-' and 'triple-blind' are sometimes used to describe these permutations. However, there is some variation in their usage and you should check to see exactly who was blinded in a trial.

Of course, it may have been impossible to blind certain groups of participants, depending on the type of intervention. Researchers should endeavour to get around this, for example by blinding outcomes assessors to the patients' treatment allocation.

Outcome measures

An outcome measure is any feature that is recorded to determine the progression of the disease or problem being studied. Outcomes should be objectively defined and measured wherever possible. Often, outcomes are expressed as mean values of measures rather than numbers of individuals having a particular outcome. The use of means can hide important information about the characteristics of patients who have improved and, perhaps more importantly, those who have got worse.

Note also that concealment of randomization, which happens before patients are enrolled, is different from blinding, which happens afterwards

Placebo control

Patients do better if they think they are receiving a treatment than if they do not. A placebo control should be use so that patients can't tell if they're on the active treatment or not.

5 Equal treatment

It should be clear from the article that, for example, there were no co-interventions which were applied to one group but not the other and that the groups were followed similarly with similar check-ups.

6 Did randomization produce comparable groups at the start of the trial?

The purpose of randomization is to generate two (or more) groups of patients who are similar in all important ways. The authors should allow you to check this by displaying important characteristics of the groups in tabular form.

Are the results important?

Two things you need to consider are how large is the treatment effect and how precise is the finding from the trial.

In any clinical therapeutic study there are three explanations for the observed effect:

1 bias;
2 chance variation between the two groups;
3 the effect of the treatment.

Could this result have happened if there was no difference between the groups?

Once bias has been excluded (by asking if the study is valid), we must consider the possibility that the results are a chance effect.

Alongside the results, the paper should report a measure of the likelihood that this result could have occurred if the treatment was no better than the control.

p values

The *p* value is a commonly used measure of this probability.

> For example, a *p* value of 0.01 means that there is a 1 in 100 (1%) probability of the result occurring by chance; $p = 0.05$ means this is a 1 in 20 probability.

Conventionally, the value of 0.05 is set as the threshold for statistical significance. If the p value is below 0.05, then the result is statistically significant; it is unlikely to have happened if there was no difference between the groups.

Confidence intervals (CIs)

> Any study can only examine a sample of a population. Hence, we would expect the sample to be different from the population. This is known as *sampling error*. Confidence intervals (CIs) are used to represent sampling error. A 95% CI specifies that there is a 95% chance that the population's 'true' value lies between the two limits.

Look to see if the confidence interval crosses the 'line of no difference' between the interventions. If so, then the result is not statistically significant.

The confidence interval is better than the *p* value because it shows you how much uncertainty there is around the stated result.

Quantifying the risk of benefit and harm

Once chance and bias have been ruled out, we must examine the difference in event rates between the control and experimental groups to see if there is a significant difference. These **event rates** can be calculated as shown below.

	Control	Experimental	Total
Event	a	b	a + b
No Event	c	d	c + d
Total	a+c	b+d	
Event rate	Control event rate CER = a/(a + c)	Experimental event rate EER = b/(b + d)	
Relative risk	EER/CER		
Absolute risk reduction	CER – EER		
Relative risk reduction	$\frac{(CER - EER)}{CER}$		

Relative risk or risk ratio (RR)
RR is the ratio of the risk in the experimental group divided by the risk in the control group.

Absolute risk reduction (ARR)
ARR is the difference between the event rates in the two groups.

Relative risk reduction (RRR)
Relative risk reduction is the ARR as a percentage of the control group risk

RR	ARR	RRR	Meaning
<1	> 0	> 0	Less events in experimental group
1	0	0	No difference between the groups
>1	< 0	< 0	More events in experimental group

ARR is a more clinically relevant measure to use than the RR or RRR. This is because relative measures 'factor out' the baseline risk, so that small differences in risk can seem significant when compared to a small baseline risk.

Number needed to treat (NNT)
Number needed to treat is the most useful measure of benefit, as it tells you the absolute number of patients who need to be treated to prevent one bad outcome. It is the inverse of the ARR:

$$\text{NNT} = 1/\text{ARR}$$

The confidence interval of an NNT is 1/the CI of its ARR:

$$95\%\ \text{CI on the ARR} = \sqrt{\frac{\text{CER}(1-\text{CER})}{n\ (\text{control})} + \frac{\text{EER}(1-\text{EER})}{n\ (\text{experimental})}}$$

Mortality in patients surviving acute myocardial infarction for at least 3 days with left ventricular ejection fraction <40% (ISIS-4, *Lancet* 1995)		Relative risk reduction (RRR)	Absolute risk reduction (ARR)	Number needed to treat (NNT)
Placebo: control event rate (CER)	Captopril: exp. event rate (EER)	$\frac{\text{CER}-\text{EER}}{\text{CER}}$	CER-EER	1/ARR
275/1116 = 0.246	228/1115 = 0.204	(0.246 – 0.204) /0.246 = 17%	0.246 – 0.204 = 0.042	1/0.042 = 24
		(NNTs always round up)		

Summary

An evidence-based approach to deciding whether a treatment is effective for your patient involves the following steps:

1. Frame the clinical question.
2. Search for evidence concerning the efficacy of the therapy.
3. Assess the methods used to carry out the trial of the therapy.
4. Determine the NNT of the therapy.
5. Decide whether the NNT can apply to your patient, and estimate a particularized NNT.
6. Incorporate your patient's values and preferences into deciding on a course of action.

Further reading

Bandolier Guide to Bias: http://www.jr2.ox.ac.uk/bandolier/band80/b80-2.html

Greenhalgh T. *How to Read a Paper*, 3rd edn. Oxford: Blackwell Publishing, 2006.

Guyatt GH, Sackett DL, Cook DJ, for the Evidence Based Medicine Working Group. Users' Guides to the Medical Literature II: How to use an article about therapy or prevention A: Are the results of the study valid? *J Am Med Assoc* 1993;**270**(21):2598–601.

Guyatt GH, Sackett DL, Cook DJ, for the Evidence Based Medicine Working Group. Users' Guides to the Medical Literature II: How to use an article