CRITICAL APPRAISAL FOR EMERGENCY MEDICINE TRAINEES
3. EVALUATION OF A THERAPY

Evaluation of a therapy involves comparing a group of patients receiving the intervention to a group of patients who do not receive it (the control group). With a few rare exceptions (such as diseases that currently have 100% mortality) a control group is always required to demonstrate that any improvement observed after treatment is not simply due to the natural course of the illness.

There are a number of key elements of the design of these studies that will determine whether the findings are valid and generalisable.

Selection and allocation of study participants
Patients are selected to a trial by a process of recruitment that usually involves identification of potential participants, assessment of eligibility using inclusion and exclusion criteria, followed by a request for consent to participate. Selection can occur at any of these stages to influence the constitution of the study population. This is obviously a necessary process in assembling the study population, but selection can influence our interpretation of the findings.

Selection of patients for a trial will clearly affect generalisability. The results of the trial will only be generalisable to patients that resemble the selected study population. If most eligible patients are identified and recruited then the results will be generalisable to the wider population. If recruitment is highly selective then findings may not be easily generalisable.

Once patients have been selected into a trial, they are then allocated to intervention or to control treatment. Bias may result if patients, carers or researchers can influence the process of allocation. For example, patients may choose a treatment that they think will be beneficial. This will result in certain types of patient being allocated to certain treatments, leading to bias. The more that patients, carers and researchers can influence allocation to treatment group, the greater bias is likely to arise. This bias may be known as allocation bias or (perhaps confusingly) selection bias.

Randomisation
Randomisation is a technique used to ensure that allocation to treatment group is not influenced by carers, patients or researchers. Patients are allocated to treatment group by a random process, such as tossing a coin. By making allocation to treatment group a random process, those involved in the trial will not be able to predict allocation, and thereby control it.

However, simply using randomisation does not eliminate allocation bias. If those involved in the trial know the randomisation schedule in advance they can select patients with a more favourable prognosis to one treatment group or another by controlling recruitment into the trial (even though they do not control allocation to treatment group). For example, we could randomise patients by randomly allocating days of the week, so that on some (random) days they receive the intervention and on others they receive control. However, if patients, carers or researchers know which treatment is being provided on that day, then they could choose participation in the

trial only if the treatment they want is being offered. Example 1 shows how this might lead to bias.

Allocation concealment
Allocation concealment ensures that a randomised trial will genuinely prevent patients, carers or researchers from influencing the allocation process. Patients, carers and researchers are not informed of the allocated treatment group for the next participant in the trial until that participant is irreversibly enrolled in the trial.

The ideal method to achieve allocation concealment is the telephone randomisation hotline. The randomisation sequence is held at a separate location that must be telephoned whenever a patient is recruited. The allocated treatment group is only revealed when all the patient's details have been recorded and they are irreversibly entered into the trial.

Consecutive, sealed, opaque envelopes can also be used to achieve allocation concealment, but all the envelopes must be accounted for at the end of the trial and regular checks must be made for tampering. It is surprising how far people will go to subvert the randomisation process!

Allocation concealment is the key to avoiding bias. Randomisation alone is not sufficient. In fact, if allocation concealment is in place then randomisation schedule does not have to be completely random. Block-randomisation, in which the randomisation sequence if split into blocks with equal (or fixed) numbers of treatments and controls, can be used to ensure equal numbers of treatment and controls in the trial. However it is important that the sequence should not be predictable, because this would mean allocation was no longer be concealed.

Allocation concealment ensures that those involved in the trial are unaware of the allocated group until the patient is irreversibly entered into the trial. Blinding refers to subsequent concealment of the treatment group from those involved in the trial. A fully blinded, placebo-controlled trial will ensure allocation concealment because (if random allocation and blinding are effective) patients, carers and researchers will be unaware of group allocation throughout the trial.

Allocation concealment applies before randomisation Blinding applies after randomisation

Blinding
Blinding tackles a different form of bias from allocation concealment. It is concerned with ensuring that the measurement of outcomes is free from bias. If any of the following individuals are aware of the treatment received they may alter their interpretation of the outcomes measured: patients, carers providing the treatment studied, carers providing subsequent care, those measuring outcomes, and those analysing outcomes. "Double blind" does not really cover the issue! If a study is described as blinded you need to identify exactly who was blinded.

The most important person to be blinded is the person measuring the outcomes. If they are aware of treatment group then results will be subject to measurement bias.

Blinding of patients and carers helps to combat the placebo effect (the beneficial effect of simply receiving treatment or attention). Whether patients or carers should be blind depends upon the type of research question. For a pragmatic trial we may simply wish to know whether treatment makes people feel better, so we are not interested in whether it is due to a placebo effect or not. For an explanatory trial we will want to know how and why the treatment is effective, so we will want to eliminate any placebo effect. This is illustrated in example 2.

In drug trials it may be possible to ensure that everyone concerned is blinded. In trials of surgery and other physical interventions this is clearly impossible. However, bias may still be minimised by ensuring that those who can be blinded are blinded. In particular, those responsible for measuring outcomes should be blind, even if carers and patients are not.

Blinding and outcomes

The potential for lack of blinding to lead to bias will depend upon the outcome being measured. "Soft" outcomes, such as patient satisfaction, quality of life, range of movement, or pain, have a strong subjective element. This does not mean that they are not important, but it does make them susceptible to bias if blinding is inadequate. "Hard" outcomes, such as death, are less subject to bias due to lack of blinding.

Intention to treat analysis (analyse as you randomise)

The main analysis should always be done on an intention to treat basis, and the overall conclusion based on this analysis. Intention to treat analysis means that patients are analysed in the group to which they were originally randomised, regardless of whether they actually received the treatment they were allocated to. It ensures that the protection from bias created by allocation concealment is maintained.

If patients are allowed to leave the group to which they were randomised this will introduce bias. Patients who withdraw, do not attend follow up, fail to comply with treatment, or have to change treatment because of adverse events will be different from those who complete their treatment as allocated. Therefore all patients should be analysed in the group they were originally allocated to, regardless of the treatment they ultimately received. Example 3 illustrates this.

Follow-up

Ideally all patients recruited into a trial should be followed-up and outcome data reported. However, this is often difficult, particularly if follow-up is prolonged and patients are mobile. If patients are lost to follow up then the researchers will have to make some sort of assumption (usually implicit) about whether those missing are typical of the study population. It is usually assumed that they are essentially similar to those followed up and do not differ between treatment groups. Therefore their loss from analysis can be accepted. If losses to follow-up are considerable (greater than 30%, for example) or differ markedly between treatment groups, this assumption is unlikely to hold, and significant bias is a reasonable possibility.

Outcome measures

There is no perfect outcome measure. "Hard" outcome measures, such as death or serious adverse events, may be resistant to bias from lack of blinding and considered "important", but are often rare and subject to type II (false negative) statistical errors.

Clinical outcomes, such as blood pressure or peak flow rate, may be sensitive to change and possible to record with adequate blinding, but are of questionable importance to the patient. Patient-centred outcomes (satisfaction, quality of life, pain) are important and relevant to the patient, but often subject to bias due to inadequate blinding. Therefore trials should ideally measure a range of different outcomes to address different objectives.

Measures of effectiveness

Rather than simply report whether a treatment is effective (i.e. is the difference in outcome between the treatment groups statistically significant?) the article should report how effective the treatment is and provide a confidence interval for this estimate. This allows the reader to decide whether the treatment effect is clinically important.

The relative risk reduction (RRR) is the difference between the intervention and control groups in the proportion of patients with the outcome (e.g. death) divided by the proportion with the outcome in the control group.

So if 20/100 patients die in the control group and 15/100 patients die in the intervention group the RRR = (0.2 – 0.15) / 0.2 = 0.25 (i.e. 25%)

The absolute risk reduction (ARR) is simply the difference between the intervention and control groups in the proportion of patients with the outcome.

So in the same example the ARR = 0.2 – 0.15 = 0.05 (i.e. 5%)

At a very simplistic level, reporting the RRR makes the treatment sound more impressive than reporting the ARR. Both measures have their uses, but the ARR may be more useful for decision-making in the individual patient, particularly if it is used to calculate the number needed to treat (NNT).

The NNT is the number of patients who would need to receive the treatment to avoid one negative outcome, such as death. It is calculated as 1 / ARR.

So in the example above the NNT = 1/0.05 = 20. We would need to treat 20 patients to avoid one death.

Summary

Critical appraisal of therapeutic evaluations is well established and based upon well-defined criteria. However, the appropriate use of selection criteria, blinding, outcome measures and follow-up will depend to a certain extent upon the aims of the trial and whether it is explanatory or pragmatic. Dogmatic application of critical appraisal checklists may lead to inappropriate rejection of useful findings.

Example 1

To determine whether acute physiotherapy is an effective intervention for soft tissue injuries presenting to the emergency department, an acute physiotherapy service was provided on randomly selected days of the week.

Randomisation of days of the week ensures that emergency department staff cannot choose which treatment to allocate patients to after they have been recruited. However, emergency department staff will know before they recruit patients whether the patient will be randomised to physiotherapy or not. Their decision to recruit patients may be influenced by awareness of which treatment they will receive. This may lead to systematic differences in the patients recruited to physiotherapy and control.

This bias could be avoided by using, for example, telephone randomisation to ensure allocation concealment. Emergency department staff would only be told whether the patient was randomised to physiotherapy or control after they have irreversibly entered the patient into the trial.

Example 2

A randomised controlled trial compared nerve block to standard analgesia for fractured neck of femur. It was decided that a placebo nerve block would not be ethical so it was not possible to blind patients to treatment group. The primary outcome was pain measured on a visual analogue scale.

Pain is a subjective experience that may be influenced by psychological factors. Patients may therefore have received some benefit from the nerve block that was not directly related to its' physiological effect. This could be described as a "placebo effect".

Interpretation of this potential bias will depend upon whether the trial had a pragmatic or explanatory aim. A pragmatic trial would simply aim to determine whether patients received better pain relief by using a nerve block, regardless of how this effect was mediated. An explanatory trial would aim to determine whether this was a physiological or a psychological effect. Thus our judgement regarding whether lack of blinding was an important flaw depends upon the aim of the trial.

Example 3

A randomised trial is undertaken to evaluate an exercise programme, outlined on a brief information sheet, for patients who have suffered an acute ankle sprain. There is no significant difference in the primary outcome measure (time to return to normal activities). However, only 55% of patients randomised to the exercise programme actually read the information sheet and followed the programme. A secondary analysis excluding the 45% who did not do the exercises shows that the programme is associated with a significant reduction in the time to return to normal activities.

This conclusion of this trial should be that there is no evidence that the exercise programme is effective. The secondary analysis is not an intention to treat analysis and carries a substantial risk of bias. Patients who do not comply with an intervention are likely to be systematically different from those who do. In the secondary analysis these patients have been excluded from the treatment group, but not the control group.