How to appraise a randomized controlled trial

Critical appraisal is an essential skill for evidence-based practitioners. It involves the systematic examination of research in order to assess its trustworthiness and relevance to specific situations. This article gives an overview on how to approach critical appraisal of randomized controlled trials.

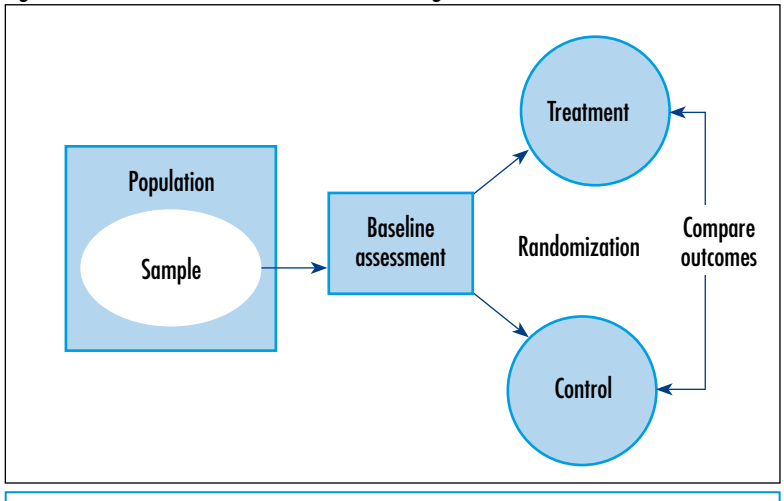
As the cultural shift to evidence-based medicine continues to embed within health-care practice, the skills required to critically appraise medical research have become highly relevant. Evidence-based medicine is defined as the 'integration of best research evidence with clinical expertise and patient values' (Sackett et al, 2000). It is often described as being based on five defined processes:

1. Asking a focused question
2. Finding the evidence
3. Critical appraisal
4. Making a decision
5. Evaluating performance (Cook et al, 1992).

This article focuses on 'critical appraisal', which aims to ensure that unreliable evidence is not translated into practice, leading to potential harm to patients and wastage of resources.

Randomized controlled trials are regarded as the gold standard for evaluating the effectiveness of interventions (Haynes et al, 2005). The key features of randomized controlled trials are that participants are randomly allocated to one or more groups to receive the intervention under evaluation or alternative intervention(s) or no intervention. The groups will therefore be similar in terms of both known and unknown confounding variables. Following randomization participants should be followed up in exactly the same way, the only difference being the intervention(s) under evaluation. This means these trials have a high level of internal validity and so any differences between the groups will be the result of the effectiveness of the intervention(s) (Torgerson and Torgerson, 2008). Thus being able to evaluate the robustness of such studies in performing such processes is vital for evidence-based practitioners. The basic structure of a parallel randomized controlled trial is illustrated in *Figure 1*. Various tools exist to aid the critical appraisal of randomized controlled trials with significant overlap (Centre for Evidence Based Medicine, 2005; Critical Appraisal Skills Programme, 2010; Higgins et al, 2011; Scottish Intercollegiate Guidelines Network, 2012); the key elements of these will be discussed in this article.

Figure 1. Parallel randomized controlled trial design.



Dr Jamilla A Hussain *is NIHR Academic Clinical Fellow in Palliative Medicine at Robert Ogden MacMillan Cancer Centre, Leeds Teaching Hospitals NHS Trust, Leeds LS9 7TF,* **Dr Amy Gadoud** *is Clinical Fellow at Hull York Medical School, University of Hull, Hull, and* **Dr Paul Taylor** *is Specialist Registrar in Palliative Medicine at St. Catherine's Hospice, Scarborough*

Correspondence to: Dr JA Hussain (jamillahussain@doctors.net.uk)

Appraising the evidence

Research literature is constantly changing and so the task of remaining up-to-date in order to practice evidence-based medicine is challenging. Although the number of studies is increasing the quality of the methods and reporting is highly variable. Critical appraisal of randomized controlled trials relies heavily on the clear and transparent reporting of the design, conduct and analysis of trials. However, reporting of randomized controlled trials is known to be incomplete and at times inaccurate (Moher et al, 2010). In view of this in 1996 the first Consolidation Standards of Reporting Trials (CONSORT) statement was developed. Subsequently two revisions (Moher et al, 2001; Schulz et al, 2010) have been published in order to keep pace with a growing evidence base. The 2010 CONSORT statement consists of a checklist of 25 essential items that should be reported in randomized controlled trial papers along with a diagram of the flow of participants through the trial (Schulz et al, 2010). Although the statement was not designed to be an appraisal tool (Moher et al, 2010), the items do help reveal any deficiencies in the research (Schulz et al, 2010).

It is also useful and efficient to use a standardized appraisal tool. The Critical Appraisal Skills Programme (2010) has developed tools for different study designs including randomized controlled trials. It describes three sequential steps to use when evaluating scientific articles:

1. Assess if the study is valid
2. What are the results?
3. Are the results applicable to my needs? (Critical Appraisal Skills Programme, 2010).

These three steps provide a useful framework for evaluating randomized controlled trials. Questions that the Critical Appraisal Skills Programme (2010) suggests will help you 'make sense of a trial' are shown in *Table 1*.

Assessing the validity of trial methodology

Did the study address a focused and appropriate research question?

The study must address an appropriate and focused research question, otherwise it may not be possible to assess the relevance of the study for any particular clinical question or if the study met its own objectives. A focused question should include four essential parts referred to as 'PICO' which identify the:

- Population
- Intervention
- Comparators
- Outcomes (Sackett et al, 1997).

Were patients randomized to treatments?

Randomization is the process where participants are assigned to the experimental or control groups with an equal probability of being assigned to any given group (Lang et al, 1997). The aim is to remove selection bias and balance confounding factors such that the control and interventional group are as similar as possible. There are several methods of randomization, and some are more prone to bias than others such as alternate allocation, allocation by date of birth or clinic day (Lang et al, 1997). All of these are not regarded as true methods of randomization and are techniques where it is easier to determine which treatment patients received (Scottish Intercollegiate Guidelines Network, 2012). Such studies are classed as controlled clinical trials not randomized controlled trials (Scottish Intercollegiate Guidelines Network, 2012). Centralized computer randomization is preferred and is often used in multicentre trials (Altman and Bland, 1999).

It is also relevant to note that the Critical Appraisal Skills Programme tool of 2006 was updated in 2010. Although the updated version no longer asks specifically 'was this a randomized controlled trial and was it appropriately so?', it remains important to assess the study design and judge whether a randomized controlled trial is the gold standard design for the research question under review, as this will not be the case for all research questions.

Was an adequate allocation concealment method used?

Allocation concealment methods are the techniques used to ensure the researchers are unaware which group participants have been allocated to until the moment of assignment. This aims to prevent researchers influencing which groups patients are allocated to. Schulz et al (1995) showed that when allocation was not concealed the estimates of treatment effect were overestimated by approximately 41% compared to those with good concealment. Third party allocation is preferred, where the investigator, on recruiting a participant, liaises with a central randomization service that allocates the trial arm (Torgerson and Roberts, 1999). If the concealment method is inadequate this introduces significant selection bias such that the results can be more harmful than observational studies that acknowledge such bias (Torgerson and Roberts, 1999).

Were participants and investigators 'blind' to treatment allocation?

Blinding is the process through which participants, health-care professionals and study personnel are prevented from knowing which treatment an individual has received when assessing the outcomes of the participants (Day and Altman, 2000). Three levels of blinding are described:

1. Single blinding: patients are unaware
2. Double blinding: patients and doctors are unaware
3. Triple blinding: patients, doctors and data analysts are unaware (Scottish Intercollegiate Guidelines Network, 2012).

The higher the level of blinding the lower the risk of bias. Sometimes, however, blinding may not be possible because of the nature of the intervention, i.e. in surgical trials, although imaginative techniques have been developed to aid the blinding process in this area (Karanicolas et al, 2010).

Table 1. Questions to help critically appraise randomized controlled trials

Was a clearly focused question asked?
Were patients randomized?
Were all patients who commenced the trial accounted for?
Were participants, staff and researchers blinded to participants' study group?
Were the groups similar at the start?
Apart from the intervention, were groups treated the same?
What were the main results?
How precise were the results?
Were all important outcomes assessed?
Can the results be generalized to the local population?
Do the benefits outweigh the costs and harms?

Adapted from Critical Appraisal Skills Programme (2010)

The extra costs and participants' burdens that may result from double or triple blinding must also be considered. For example when the two treatments in a randomized controlled trial cannot be made identical, a double dummy is required and the subjects are required to take two sets of treatment, one active and one placebo, rather than just one. The potential effect of not blinding on the specific outcomes measured should therefore be taken into account before burdening patients. For instance if the outcome was cancer-free survival this is not likely to be affected by not blinding, whereas pain intensity and quality of life may be. Therefore blinding and level of blinding needs to be judged in relation to the study outcome measures.

Were the treatment and control groups similar at the start?

The randomization process should ensure that the different groups are comparable with regards to prognostic factors (Scottish Intercollegiate Guidelines Network, 2012). Any statistically significant differences in important demographic and comorbidity factors such as age, gender, social class, ethnicity and disease stage should be reported.

Was the only difference between the groups the intervention under investigation?

Besides the intervention under consideration both groups should be treated equally. If some patients receive additional treatment bias is potentially introduced.

Were all participants accounted for?

Studies must account for all participants who entered the trial. The amount of withdrawal from a study should be noted as this could bias results. Traditionally, a 20% dropout rate is considered acceptable and is usually adjusted for in the power calculation (Kadam and Bhalerao, 2010). As well as determining how many patients withdrew it is important to consider why they withdrew and if there is differential dropout in the treatment arms.

Was an intention to treat analysis performed?

Intention to treat analysis is the analysis of all participant data according to the group they were randomized to, rather than the actual intervention received (Gupta, 2011). This analysis technique preserves the benefits of randomization and more closely mirrors effectiveness in clinical practice where full concordance with treatment is unlikely. Strict intention to treat analysis is difficult to achieve, however, because of missing data and non-adherence to trial protocol (Moher et al, 2010).

Patients with missing data can either be excluded from the analysis (complete case analysis) or the missing outcomes can be imputed (Altman, 2009). Removing patients with missing data introduces bias if the withdrawal was related to the intervention and additionally there is a reduction in the sample size leading to loss of power. Thus the analysis is no longer strictly intention to treat. Missing data can be estimated from other collected data, but this relies on strong assumptions about the data (Shih, 2002).

Some participants may not have completed the protocol fully; in this situation the options are either to exclude the patients or conduct a 'per-protocol' analysis, which excludes participants who did not receive a defined minimum amount of intervention. Both of these again deviate from the CONSORT principles of intention to treat analysis and could result in significant bias (Moher et al, 2010).

In view of these difficulties the CONSORT checklist (Moher et al, 2010) recommends that a clear description of who was included in each analysis is essential in order to assess such biases.

What are the results?

If the methodology is considered adequate the next step is to assess the reliability of the results.

What was the treatment effect estimate?

The treatment effect of an intervention can be reported in several formats including absolute risk, relative risk, odds ratio and number needed to treat. These are statistical estimates and should be reported with confidence intervals (Centre for Evidence Based Medicine, 2005). An explanation of the terms and calculations can be found in *Table 2*.

Absolute risk measures (absolute risk reduction and number needed to treat), unlike relative risk measures, take account of the baseline risk and so are regarded as being better at discriminating between small and large treatment effects (Sackett et al, 2000). Although the effect size and statistical significance are important, an assessment of clinical significance is also required (Man-Son-Hing et al, 2002). Clinical significance is the difference in effect size between groups that is considered important to clinical decision making, and does not necessarily require the difference to be statistically significant. It is important to remember that 'not statistically significant' does not mean 'no difference'; it simply means there is insufficient evidence for a difference. This is where clinical experience and patient values, as described in the definition of evidence-based medicine at the start, are important.

How precise was the estimate of treatment effect?

Based on the sample risk the population risk of an outcome can be estimated; this is known as the point estimate. How close this estimate is to the population value is assessed by the confidence intervals of the estimate. Confidence intervals are used to assess the statistical

significance and the precision of a treatment effect (Gardner and Altman, 1986). A wide confidence interval width indicates less precision. If the value corresponding to no effect (the null hypothesis) is contained within the 95% confidence interval then the result is classed as 'not significant' at the 5% significance level. If the value corresponding to no effect is outside the 95% confidence interval the result is 'significant' at the 5% significance level.

Are the results applicable to my needs?

Before incorporating evidence into clinical practice it is important for practitioners to consider the following questions that relate to the external validity of the results (Critical Appraisal Skills Programme, 2010):

Can the results be applied to the local population?

It is important to assess the similarity and differences between the local and study populations. Clinical judgement is required to establish what is a relevant difference and should take into account published evidence regarding the likely relevance of the noted variables. Important factors to consider include biology, geography and culture (Akobeng, 2005). A useful tool is to think about whether the patient under consideration for the intervention would have been eligible to enter the study.

Were all clinically important outcomes considered?

Not only is clinical effectiveness an important consideration, but also cost-effectiveness, convenience, patient-values and the expertise required. These all need to be considered and discussed before changing practice.

Do the benefits outweigh the costs and potential harms?

When considering a new intervention it is important to consider the treatment with the patient in mind. For each patient clinicians must weigh up the potential side effects and comorbidities which may affect the risk/benefit balance. Using clinical judgment and patient preference is important when making such decisions.

Conclusions

Randomized controlled trials that are well designed and conducted provide the most reliable evaluation of the effectiveness of health-care interventions. Trials that are inadequately conducted and reported, however, could result in significant bias and exaggerated treatment effects. It is important for medical practitioners to be able to critically appraise randomized controlled trials in order to prevent such biases misleading clinical decisions. Using the three steps outlined by the Critical Appraisal Skills Programme provides a useful framework for this process. BJHM

Table 2. Treatment effect estimates

For example in a randomized controlled trial where 10% of the control group and 5% of the intervention group die at 1 year:	
Absolute risk reduction	Arithmetic difference between the absolute risk of an outcome in individuals who were exposed against those who were not exposed (10% – 5% = 5%)
Number needed to treat	The reciprocal of absolute risk reduction. It is the number of people who need to receive an intervention in order to achieve the required outcome in one of them. Number needed to treat is regarded as the clinically useful way to describe risk (1/0.05 = 20)
Relative risk	The ratio of an outcome in the exposed group to the risk of the outcome in the unexposed group (0.05/0.10 =0.5)
Relative risk reduction	This measures how much the risk is reduced in the treatment group compared to the control group (= 1 – relative risk = 1 – 0.5 = 0.5)

Conflict of interest: none.

Akobeng AK (2005) Understanding measures of treatment effect in clinical trials. *Arch Dis Child* **90**: 54–6

Altman DG (2009) Missing outcomes in randomised trials: addressing the dilemma. *Open Med* **3**: e21–3

Altman DG, Bland JM (1999) How to randomise. *BMJ* **319**(7211): 703–4

Centre for Evidence Based Medicine (2005) RCT Critical Appraisal Sheet. www.cebm.net/index.aspx?o=1097 (accessed 25 March 2013)

Cook DJ, Jaeschke R, Guyatt GH (1992) Critical appraisal of therapeutic interventions in the intensive care unit: human monoclonal antibody treatment in sepsis. Journal Club of the Hamilton Regional Critical Care Group. *J Intensive Care Med* **7**(6): 275–82

Critical Appraisal Skills Programme (2010) Making sense of evidence about clinical effectiveness. www.casp-uk.net/wp-content/uploads/2011/11/CASP_RCT_Appraisal_Checklist_14oct10.pdf (accessed 21 March 2013)

Day SJ, Altman DG (2000) Blinding in clinical trials and other studies. *BMJ* **321**: 504

Gardner MJ, Altman DG (1986) Confidence intervals rather than P values: estimation rather than hypothesis testing. *BMJ* **292**(6522): 746–50

Gupta SK (2011) Intention-to-treat concept: A review. *Perspect Clin Res* **2**(3): 109–12

Haynes RB, Sackett DL, Guyatt GH, Tugwell P (2005) *Clinical Epidemiology: How to Do Clinical Practice Research.* 2nd edn. Lippincott Williams & Wilkins, London

Higgins JPT, Altman DG, Sterne JAC, eds (2011) Assessing risk of bias in included studies. In: Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. www.cochrane-handbook.org (accessed 25 March 2013)

KEY POINTS

- Critical appraisal is a key step in practicing evidence-based medicine.
- Randomized controlled trials are regarded as the gold standard in the evaluation of effectiveness of interventions.
- Several standardized tools exist to help practitioners critically appraise randomized controlled trials.
- Three steps to evaluating randomized controlled trials are suggested: first to assess if the study is valid, then to establish what the results are and finally to confirm whether the results are applicable to your particular clinical context.

Kadam P, Bhalerao S (2010) Sample size calculation. *Int J Ayurdeva Res* **1**(1): 55–7
Karanicolas PJ, Farrokhyar F, Bhandari M (2010) Blinding: Who, what, when, why, how? *Can J Surg* **53**(5): 345–8
Lang TA, Secic M (1997) *How to Report Statistics in Medicine.* American College of Physicians, Philadelphia
Man-Son-Hing M, Laupacis A, O'Rourke K, Molnar FJ, Mahon J, Chan KBY, Wells G (2002) Determination of the clinical importance of study results: a review. *J Gen Intern Med* **17**(6): 469–76
Moher D, Schulz KF, Altman DG, for the CONSORT Group (2001) The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomised trials. *Ann Intern Med* **134**: 657–62
Moher D, Hopewell S, Schulz KF et al (2010) CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. *J Clin Epidemiol* **63**(8): e1–37
Sackett DL, Richardson WS, Rosenberg W, Haynes RB (1997) *Evidence-based medicine: How to practice and teach EBM.* 1st edn. Churchill Livingstone, London
Sackett DL, Strauss SE, Richardson WS et al (2000) *Evidence-based medicine: how to practice and teach EBM.* 2nd edn. Churchill Livingstone, London
Schulz KF, Chalmers I, Hayes RJ et al (1995) Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* **273**: 408–12
Schulz KF, Altman DG, Moher D, for the CONSORT Group (2010) CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ* **340**: c332
Scottish Intercollegiate Guidelines Network (2012) Methodology Checklist 2: Randomised Controlled Trials. www.sign.ac.uk/methodology/checklists/20121206_Checklist_for_controlled_trials.doc (accessed 25 March 2013)
Shih WJ (2002) Problems in dealing with missing data and informative censoring in clinical trials. *Curr Control Trials Cardiovasc Med* **3**(1): 4
Torgerson DJ, Roberts C (1999) Randomisation methods: concealment. *BMJ* **319**: 375–6
Torgerson DJ, Torgerson CJ (2008) *Designing randomised trials in health, education, and the social sciences: An introduction.* Palgrave Macmillan, Hampshire