REFERENCES

1 **Sackett DL**, Strauss SE, Richardson WS, *et al. Evidence-based medicine: how to practice and teach EBM.* London: Churchill-Livingstone, 2000.
2 **Steves R**, Hootman JM. Evidence-based medicine: what is it and how does it apply to athletic training? *J Athl Train* 2004;**39**:83–7.
3 **Brownson RC**, Baker EA, Leet TL, *et al. Evidence based public Health.* New York: Oxford University Press, 2003.
4 **Gray GE**, Pinson LA. Evidence-based medicine and psychiatric practice. *Psychiatr Q* 2003;**74**:387–99.
5 **Sackett DL**. Evidence-based medicine. *Semin Perinatol* 1997;**21**:3–5.
6 **Levi M**. Formulating clinical questions. In: McGovern DPB, Valori RM, Summerskill WSM, Levi M, eds. *Key topics in evidence based medicine.* Oxford; BIOS Scientific Publishers, 2001.
7 **Carneiro AV**. The correct formulation of clinical questions for the practice of evidence based medicine. *Acta Med Port* 1998;**11**:745–8.
8 **Rosenberg WM**, Sackett DL. On the need for evidence-based medicine. *Therapie* 1996;**51**:212–7.
9 **Jordaan M**, Jones R. Adoption of internet technology by UK postgraduate centres: a questionnaire survey. *Health Libr Rev* 1999;**16**:166–73.
10 **Rosenberg WM**, Deeks J, Lusher A, *et al.* Improving searching skills and evidence retrieval. *J R Coll Physicians Lond* 1998;**32**:557–63.
11 **Craig JV**, Smyth RL. *The evidence-based manual for nurses.* London: Churchill Livingstone, 2002.
12 **Patel H**, Platt R, Lozano JM, *et al.* Glucocorticoids for acute viral bronchiolitis in infants and young children. *The Cochrane database of systematic reviews,* 2004, issue 3.
13 **Rosenberg W**, Donald A. Evidence based medicine: an approach to clinical problem-solving. *BMJ* 1995;**310**:1122–6.
14 **Critical Appraisal Skills Programme**. Appraisal Tools. Oxford, UK. http://www.phru.nhs.uk/casp/appraisa.htm (accessed 10 Dec 2004).
15 **Straus SE**, Sackett DL. Using research findings in clinical practice. *BMJ* 1998;**317**:339–42.

EVIDENCE BASED CHILD HEALTH 2

Understanding randomised controlled trials

A K Akobeng

The hierarchy of evidence in assessing the effectiveness of interventions or treatments is explained, and the gold standard for evaluating the effectiveness of interventions, the randomised controlled trial, is discussed. Issues that need to be considered during the critical appraisal of randomised controlled trials, such as assessing the validity of trial methodology and the magnitude and precision of the treatment effect, and deciding on the applicability of research results, are discussed. Important terminologies such as randomisation, allocation concealment, blinding, intention to treat, p values, and confidence intervals are explained.

In the first article of the series,[1] I described evidence based medicine (EBM) as a systematic approach to clinical problem solving, which allows the integration of the best available research evidence with clinical expertise and patient values. In this article, I will explain the hierarchy of evidence in assessing the effectiveness of interventions or treatments, and discuss the randomised controlled trial, the gold standard for evaluating the effectiveness of interventions.

HIERARCHY OF EVIDENCE

It is well recognised that some research designs are more powerful than others in their ability to answer research questions on the effectiveness of interventions. This notion has given rise to the concept of "hierarchy of evidence". The hierarchy provides a framework for ranking evidence that evaluates health care interventions and indicates which studies should be given most weight in an evaluation where the same question has been examined using different types of study.[2]

Figure 1 illustrates such a hierarchy. The ranking has an evolutionary order, moving from simple observational methods at the bottom, through to increasingly rigorous methodologies. The pyramid shape is used to illustrate the increasing risk of bias inherent in study designs as one goes down the pyramid.[3] The randomised controlled trial (RCT) is considered to provide the most reliable evidence on the effectiveness of interventions because the processes used during the conduct of an RCT minimise the risk of confounding factors influencing the results. Because of this, the findings generated by RCTs are likely to be closer to the true effect than the findings generated by other research methods.[4]

The hierarchy implies that when we are looking for evidence on the effectiveness of interventions or treatments, properly conducted systematic reviews of RCTs with or without meta-analysis or properly conducted RCTs will provide the most powerful form of evidence.[3] For example, if you want to know whether there is good evidence that children with meningitis should be given corticosteroids or not, the best articles to look for would be systematic reviews or RCTs.

WHAT IS A RANDOMISED CONTROLLED TRIAL?

An RCT is a type of study in which participants are randomly assigned to one of two or more clinical interventions. The RCT is the most scientifically rigorous method of hypothesis testing available,[5] and is regarded as the gold standard trial for evaluating the effectiveness of interventions.[6] The basic structure of an RCT is shown in fig 2.

Correspondence to:
Dr A K Akobeng,
Department of Paediatric Gastroenterology, Central Manchester and Manchester Children's University Hospitals, Booth Hall Children's Hospital, Charlestown Road, Blackley, Manchester, M9 7AA, UK; tony.akobeng@cmmc.nhs.uk

Accepted 22 April 2005

Abbreviations: Abbreviations: CONSORT, consolidated standards of reporting trials; EBM, evidence based medicine; PCDAI, paediatric Crohn's disease activity index; RCT, randomised controlled trial
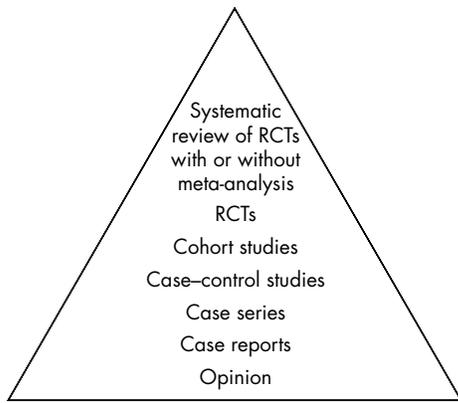
Figure 1 Hierarchy of evidence for questions about the effectiveness of an intervention or treatment.

Figure 2 The basic structure of a randomised controlled trial.

A sample of the population of interest is randomly allocated to one or another intervention and the two groups are followed up for
a specified period of time. Apart from the interventions being compared, the two groups are treated and observed in an identical manner. At the end of the study, the groups are analysed in terms of outcomes defined at the outset. The results from, say, the treatment A group are compared with results from the treatment B group. As the groups are treated identically apart from the intervention received, any differences in outcomes are attributed to the trial therapy.[6]

WHY A RANDOMISED CONTROLLED TRIAL?

The main purpose of random assignment is to prevent selection bias by distributing the characteristics of patients that may influence the outcome randomly between the groups, so that any difference in outcome can be explained only by the treatment.[7] Thus random allocation makes it more likely that there will be balancing of baseline systematic differences between intervention groups with regard to known and unknown factors—such as age, sex, disease activity, and duration of disease—that may affect the outcome.

APPRAISING A RANDOMISED CONTROLLED TRIAL

When you are reading an RCT article, the answers to a few questions will help you decide whether you can trust the results of the study and whether you can apply the results to your patient or population. Issues to consider when reading an RCT may be condensed into three important areas[8]:

- the validity of the trial methodology;
- the magnitude and precision of the treatment effect;
- the applicability of the results to your patient or population.

A list of 10 questions that may be used for critical appraisal of an RCT in all three areas is given in box 1.[9]

ASSESSING THE VALIDITY OF TRIAL METHODOLOGY
Focused research question

It is important that research questions be clearly defined at the outset. The question should be focused on the problem of interest, and should be framed in such a way that even somebody who is not a specialist in the field would understand why the study was undertaken.
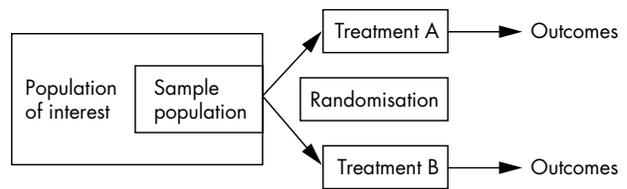
Randomisation

Randomisation refers to the process of assigning study participants to experimental or control groups at random such that each participant has an equal probability of being assigned to any given group.[10] The main purpose of randomisation is to eliminate selection bias and balance known and unknown confounding factors in order to create a control group that is as similar as possible to the treatment group.

Methods for randomly assigning participants to groups, which limits bias, include the use of a table of random numbers and a computer program that generates random numbers. Methods of assignment that are prone to bias include alternating assignment or assignment by date of birth or hospital admission number.[10]

In very large clinical trials, simple randomisation may lead to a balance between groups in the number of patients allocated to each of the groups, and in patient characteristics. However, in ''smaller'' studies this may not be the case. Block randomisation and stratification are strategies that may be used to help ensure balance between groups in size and patient characteristics.[11]

Block randomisation

Block randomisation may be used to ensure a balance in the number of patients allocated to each of the groups in the trial. Participants are considered in blocks of, say, four at a time. Using a block size of four for two treatment arms (A and B) will lead to six possible arrangements of two As and two Bs (blocks):
AABB BBAA ABAB BABA ABBA BAAB
A random number sequence is used to select a particular block, which determines the allocation order for the first four

---

Box 1: Questions to consider when assessing an RCT[9]

- Did the study ask a clearly focused question?
- Was the study an RCT and was it appropriately so?
- Were participants appropriately allocated to intervention and control groups?
- Were participants, staff, and study personnel blind to participants' study groups?
- Were all the participants who entered the trial accounted for at its conclusion?
- Were participants in all groups followed up and data collected in the same way?
- Did the study have enough participants to minimise the play of chance?
- How are the results presented and what are the main results?
- How precise are the results?
- Were all important outcomes considered and can the results be applied to your local population?

subjects. In the same vein, treatment group is allocated to the next four patients in the order specified by the next randomly selected block.

Stratification

While randomisation may help remove selection bias, it does not always guarantee that the groups will be similar with regard to important patient characteristics.[12] In many studies, important prognostic factors are known before the study. One way of trying to ensure that the groups are as identical as possible is to generate separate block randomisation lists for different combinations of prognostic factors. This method is called stratification or stratified block sampling. For example, in a trial of enteral nutrition in the induction of remission in active Crohn's disease, potential stratification factors might be disease activity (paediatric Crohn's disease activity index (PCDAI) $\leqslant 25$ v >25) and disease location (small bowel involvement v no small bowel involvement). A set of blocks could be generated for those patients who have PCDAI $\leqslant 25$ and have small bowel disease; those who have PCDAI $\leqslant 25$ and have no small bowel disease; those who have PCDAI >25 and have small bowel disease; and those who have PCDAI >25 and have no small bowel disease.

Allocation concealment

Allocation concealment is a technique that is used to help prevent selection bias by concealing the allocation sequence from those assigning participants to intervention groups, until the moment of assignment. The technique prevents researchers from consciously or unconsciously influencing which participants are assigned to a given intervention group. For instance, if the randomisation sequence shows that patient number 9 will receive treatment A, allocation concealment will remove the ability of researchers or other health care professionals from manoeuvring to place another patient in position 9.

In a recent observational study, Schulz et al showed that in trials in which allocation was not concealed, estimates of treatment effect were exaggerated by about 41% compared with those that reported adequate allocation concealment.[13]

A common way for concealing allocation is to seal each individual assignment in an opaque envelope.[10] However, this method may have disadvantages, and "distance" randomisation is generally preferred.[14] Distance randomisation means that assignment sequence should be completely removed from those who make the assignments. The investigator, on recruiting a patient, telephones a central randomisation service which issues the treatment allocation.

Although an RCT should, in theory, eliminate selection bias, there are instances where bias can occur.[15] You should not assume that a trial methodology is valid merely because it is stated to be an RCT. Any selection bias in an RCT invalidates the study design and makes the results no more reliable than an observational study. As Torgesson and Roberts have suggested, the results of a supposed RCT which has had its randomisation compromised by, say, poor allocation concealment may be more damaging than an explicitly unrandomised study, as bias in the latter is acknowledged and the statistical analysis and subsequent interpretation might have taken this into account.[14]

Blinding

There is always a risk in clinical trials that perceptions about the advantages of one treatment over another might influence outcomes, leading to biased results. This is particularly important when subjective outcome measures are being used. Patients who are aware that they are receiving what they believe to be an expensive new treatment may report being better than they really are. The judgement of a doctor who expects a particular treatment to be more effective than another may be clouded in favour of what he perceives to be the more effective treatment. When people analysing data know which treatment group was which, there can be the tendency to "overanalyse" the data for any minor differences that would support one treatment.

Knowledge of treatment received could also influence management of patients during the trial, and this can be a source of bias. For example, there could be the temptation for a doctor to give more care and attention during the study to patients receiving what he perceives to be the less effective treatment in order to compensate for perceived disadvantages.

To control for these biases, "blinding" may be undertaken. The term blinding (sometimes called masking) refers to the practice of preventing study participants, health care professionals, and those collecting and analysing data from knowing who is in the experimental group and who is in the control group, in order to avoid them being influenced by such knowledge.[16] It is important for authors of papers describing RCTs to state clearly whether participants, researchers, or data evaluators were or were not aware of assigned treatment.

In a study where participants do not know the details of the treatment but the researchers do, the term "single blind" is used. When both participants and data collectors (health care professionals, investigators) are kept ignorant of the assigned treatment, the term "double blind" is used. When, rarely, study participants, data collectors, and data evaluators such as statisticians are all blinded, the study is referred to as "triple blind".[5]

Recent studies have shown that blinding of patients and health care professionals prevents bias. Trials that were not double blinded yielded larger estimates of treatment effects than trials in which authors reported double blinding (odds ratios exaggerated, on average, by 17%).[17]

It should be noted that, although blinding helps prevent bias, its effect in doing so is weaker than that of allocation concealment.[17] Moreover, unlike allocation concealment, blinding is not always appropriate or possible. For example, in a randomised controlled trial where one is comparing enteral nutrition with corticosteroids in the treatment of children with active Crohn's disease, it may be impossible to blind participants and health care professionals to assigned intervention, although it may still be possible to blind those analysing the data, such as statisticians.

Intention to treat analysis

As stated earlier, the validity of an RCT depends greatly on the randomisation process. Randomisation ensures that known and unknown baseline confounding factors would balance out in the treatment and control groups. However, after randomisation, it is almost inevitable that some participants would not complete the study for whatever reason. Participants may deviate from the intended protocol because of misdiagnosis, non-compliance, or withdrawal. When such patients are excluded from the analysis, we can no longer be sure that important baseline prognostic factors in the two groups are similar. Thus the main rationale for random allocation is defeated, leading to potential bias.

To reduce this bias, results should be analysed on an "intention to treat" basis.

Intention to treat analysis is a strategy in the conduct and analysis of randomised controlled trials that ensures that all patients allocated to either the treatment or control groups are analysed together as representing that treatment arm whether or not they received the prescribed treatment or completed the study.[5] Intention to treat introduces clinical reality into research by recognising that for several reasons, not all participants randomised will receive the intended treatment or complete the follow up.[18]

According to the revised CONSORT statement for reporting RCTs, authors of papers should state clearly which participants are included in their analyses.[19] The sample size per group, or the denominator when proportions are being reported, should be provided for all summary information. The main results should be analysed on the basis of intention to treat. Where necessary, additional analyses restricted only to participants who fulfilled the intended protocol (per protocol analyses) may also be reported.

Power and sample size calculation

The statistical power of an RCT is the ability of the study to detect a difference between the groups when such a difference exists. The power of a study is determined by several factors, including the frequency of the outcome being studied, the magnitude of the effect, the study design, and the sample size.[5] For an RCT to have a reasonable chance of answering the research question it addresses, the sample size must be large enough—that is, there must be enough participants in each group.

When the sample size of a study is too small, it may be impossible to detect any true differences in outcome between the groups. Such a study might be a waste of resources and potentially unethical. Frequently, however, small sized studies are published that claim no difference in outcome between groups without reporting the power of the studies. Researchers should ensure at the planning stage that there are enough participants to ensure that the study has a high probability of detecting as statistically significant the smallest effect that would be regarded as clinically important.[20]

MAGNITUDE AND SIGNIFICANCE OF TREATMENT EFFECT

Once you have decided that the methodology of a study is valid within reason, the next step is to decide whether the results are reliable. Two things usually come into mind in making this decision—how big is the treatment effect, and how likely is it that the result obtained is due to chance alone?

Magnitude of treatment effect

Magnitude refers to the size of the measure of effect. Treatment effect in RCTs may be reported in various ways including absolute risk, relative risk, odds ratio, and number needed to treat. These measures of treatment effect and their advantages and disadvantages have recently been reviewed.[21] A large treatment effect may be more important than a small one.

Statistical significance

Statistical significance refers to the likelihood that the results obtained in a study were not due to chance alone. Probability (p) values and confidence intervals may be used to assess statistical significance.

p Value

A p value can be thought of as the probability that the observed difference between two treatment groups might have occurred by chance. The choice of a significance level is artificial but by convention, many researchers use a p value of 0.05 as the cut off for significance. What this means is that if the p value is less than 0.05, the observed difference between the groups is so unlikely to have occurred by chance that we reject the null hypothesis (that there is no difference) and accept the alternative hypothesis that there is a real difference between the treatment groups. When the p value is below the chosen cut off, say 0.05, the result is generally referred to as being statistically significant. If the p value is greater than 0.05, then we say that the observed difference might have occurred by chance and we fail to reject the null

hypothesis. In such a situation, we are unable to demonstrate a difference between the groups and the result is usually referred to as not statistically significant.

Confidence intervals

The results of any study are estimates of what might happen if the treatment were to be given to the entire population of interest. When I test a new asthma drug on a randomly selected sample of children with asthma in the United Kingdom, the treatment effect I will get will be an estimate of the "true" treatment effect for the whole population of children with asthma in the country. The 95% confidence interval (CI) of the estimate will be the range within which we are 95% certain that the true population treatment effect will lie. It is most common to report 95% CI, but other intervals, such as 90% and 99% CI, may also be calculated for an estimate.

If the CI for a mean difference *includes* 0, then we have been unable to demonstrate a difference between the groups being compared ("not statistically significant"), but if the CI for a mean difference does *not* include 0, then a statistically significant difference between the groups has been shown. In the same vein, if the CI for relative risk or odds ratio for an estimate includes 1, then we have been unable to demonstrate a statistically significant difference between the groups being compared, and if it does not include 1, then there is a statistically significant difference.

Confidence intervals versus p values

CIs convey more useful information than p values. CI may be used to assess statistical significance, provide a range of plausible values for a population parameter, and gives an idea about how precise the measured treatment effect is (see below). Authors of articles could report both p values and CIs.[22] However, if only one is to be reported, then it should be the CI, as the p value is less important and can be deduced from the CI; p values tell us little extra when CIs are known.[22 23]

Clinical significance

A statistically significant finding by itself can have very little to do with clinical practice and has no direct relation to clinical significance. Clinical significance reflects the value of the results to patients and may be defined as a difference in effect size between groups that could be considered to be important in clinical decision making, regardless of whether the difference is statistically significant or not. Magnitude and statistical significance are numerical calculations, but judgements about the clinical significance or clinical importance of the measured effect are relative to the topic of interest.[2] Judgements about clinical significance should take into consideration how the benefits and any adverse events of an intervention are valued by the patient.

PRECISION OF TREATMENT EFFECT

CI is important because it gives an idea about how precise an estimate is. The width of the interval indicates the precision of the estimate. The wider the interval, the less the precision. A very wide interval may indicate that more data should be collected before anything definite can be said about the estimate.

APPLYING RESULTS TO YOUR OWN PATIENTS

An important concept of EBM is that clinicians should make decisions about whether the valid results of a study are applicable to their patients. The fact that good evidence is available on a particular asthma treatment does not necessarily mean that all patients with asthma can or should be given that treatment. Some of the issues one needs to consider before deciding whether to incorporate a particular

piece of research evidence into clinical practice are briefly discussed below.

Are the participants in the study similar enough to my patients?

If a particular drug has been found to be effective in adults with meningitis in the USA, you need to decide whether there is any biological, geographical, or cultural reason why that particular drug will not be effective in children with meningitis in the United Kingdom.

Do the potential side effects of the drug outweigh the benefits?

If a particular treatment is found to be effective in an RCT, you need to consider whether the reported or known side effects of the drug may outweigh its potential benefits to your patient. You may also need to consider whether an individual patient has any potential co-morbid condition which may alter the balance of benefits and risks. In such a situation, you may, after consultation with the patient or carers, decide not to offer the treatment.

Does the treatment conflict with the patient's values and expectations?

Full information about the treatment should be given to the patient or carers, and their views on the treatment should be taken into account. A judgement should be made about how the patient and carers value the potential benefits of the treatment as against potential harms.

Is the treatment available and is my hospital prepared to fund it?

There will be no point in prescribing a treatment which cannot either be obtained in your area of work or which your hospital or practice is not in a position to fund, for whatever reason, including cost.

CONCLUSIONS

An RCT is the most rigorous scientific method for evaluating the effectiveness of health care interventions. However, bias could arise when there are flaws in the design and management of a trial. It is important for people reading medical reports to develop the skills for critically appraising RCTs, including the ability to assess the validity of trial methodology, the magnitude and precision of the treatment effect, and the applicability of results.

Competing interests: none declared

REFERENCES

1 **Akobeng AK**. Evidence-based child health. 1. Principles of evidence-based medicine. *Arch Dis Child* 2005;**90**:837–40.
2 **Rychetnik L**, Hawe P, Waters E, *et al.* A glossary for evidence based public health. *J Epidemiol Community Health* 2004;**58**:538–45.
3 **Craig JV**, Smyth RL. *The evidence-based manual for nurses*. London: Churchill Livingstone, 2002.
4 **Evans D**. Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. *J Clin Nurs* 2003;**12**:77–84.
5 **Last JM**. *A dictionary of epidemiology*. New York: Oxford University Press, 2001.
6 **McGovern DPB**. Randomized controlled trials. In: McGovern DPB, Valori RM, Summerskill WSM, eds. *Key topics in evidence based medicine*. Oxford: BIOS Scientific Publishers, 2001:26–9.
7 **Roberts C**, Torgesson D. Randomisation methods in controlled trials. *BMJ* 1998;**317**:1301–10.
8 **Sackett DL**, Strauss SE, Richardson WS, *et al. evidence-based medicine: how to practice and teach EBM*. London: Churchill-Livingstone, 2000.
9 **Critical Appraisal Skills Programme**. *Appraisal tools*. Oxford, UK, http://www.phru.nhs.uk/casp/rcts.htm (accessed 8 December 2004).
10 **Lang TA**, Secic M. *How to report statistics in medicine*. Philadelphia: American College of Physicians, 1997.
11 **Altman DG**, Bland JM. How to randomise. *BMJ* 1999;**319**:703–4.
12 **Chia KS**. Randomisation: magical cure for bias. *Ann Acad Med Singapore* 2000;**29**:563–4.
13 **Schulz KF**, Chalmers I, Hayes RJ, *et al.* Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;**273**:408–12.
14 **Torgerson DJ**, Roberts C. Randomisation methods: concealment. *BMJ* 1999;**319**:375–6.
15 **Torgerson DJ**, Torgerson CJ. Avoiding bias in randomised controlled trials in educational research. *Br J Educ Stud* 2003;**51**:36–45.
16 **Day SJ**, Altman DG. Blinding in clinical trials and other studies. *BMJ* 2000;**321**:504.
17 **Schulz KF**. Assessing allocation concealment and blinding in randomised controlled trials: why bother? *Evid Based Nurs* 2000;**5**:36–7.
18 **Summerskill WSM**. Intention to treat. In: McGovern DPB, Valori RM, Summerskill WSM, eds. *Key topics in evidence based medicine*. Oxford: BIOS Scientific Publishers, 2001:105–7.
19 **Altman DG**, Schulz KF, Moher D, *et al.* The revised CONSORT statement for reporting randomised controlled trials: explanation and elaboration. *Ann Intern Med* 2001;**134**:663–94.
20 **Devane D**, Begley CM, Clarke M. How many do I need? Basic principles of sample size estimation. *J Adv Nurs* 2004;**47**:297–302.
21 **Akobeng AK**. Understanding measures of treatment effect in clinical trials. *Arch Dis Child*, 2005;90, 54–6.
22 **Altman DG**. *Practical statistics for medical research*. London: Chapman and Hall/CRC, 1991:152–78.
23 **Coggon D**. *Statistics in clinical practice*. London: BMJ Publishing Group, 1995.