

Randomized clinical trials

One pill makes you larger,
and one pill makes you small.
And the ones your mother gives you,
don't do anything at all.

Grace Slick, The Jefferson Airplane: White Rabbit, from Surrealistic Pillow, 1967

Learning objectives

In this chapter you will learn:

- the unique features of randomized clinical trials (RCTs)
- how to undertake critical interpretation of RCTs

The randomized clinical trial (RCT) is the ultimate paradigm of clinical research. Many consider the RCT to be the most important medical development of the twentieth century, as their results are used to dictate clinical practice. Although these trials are often put on a pedestal, it is important to realize that as with all experiments, there may be flaws in the design, implementation, and interpretation of these trials. The competent reader of the medical literature should be able to evaluate the results of a clinical trial in the context of the potential biases introduced into the research experiment, and determine if it contains any fatal flaws

Introduction

The clinical trial is a relatively recent development in medical research. Prior to the 1950s, most research was based upon case series or uncontrolled observations. James Lind, a surgeon in the

British Navy, can claim credit for performing the first recorded clinical trial. In 1747, aboard the ship *Salisbury*, he took 12 sailors with scurvy and divided them into six groups of two each. He made sure they were similar in every way except for the treatment they received for scurvy. Dr. Lind found that the two sailors who were given oranges and lemons got better while the other ten did not. After that trial, the process of the clinical trial went relatively unused until it was revived with studies of the efficacy of streptomycin for the treatment of tuberculosis done in 1950. The randomized clinical trial or randomized controlled trial has remained the premier source of new knowledge in medicine since then.

A **randomized clinical trial** is an experiment. In an RCT, subjects are randomly assigned to one of two or more therapies and then treated in an identical manner for all other potential variables. . Subjects in an RCT are just as likely as unlikely to get the therapy of interest as they are to get the comparator therapy. Ideally the researchers are blinded to the group in which the subjects are allocated. The randomization code is not broken until the study is finally completed. There are variations on this theme using blinded safety committees to determine if the study should be stopped. Sometimes it is warranted to release the results of the study, which is stopped early because it showed a huge benefit and continuing the study would not be ethical.

Physician decision making and RCTs

There are several ways that physicians make decisions on the best treatment for their patients.

Induction is the retrospective analysis of uncontrolled clinical experience or extension of the expected mechanism of disease as taught in pathophysiology. It is doing that which “seems to work,” “worked before,” or “ought to work.” **Abdication** or seduction is someone doing something because others tell them that it is the right thing to do. These may be teachers, consultants, colleagues, advertisements, pharmaceutical representatives, authors of medical

textbooks, and others. One accepts their analysis of the medical information on faith and this dictates what one actually does for his or her patient.

Deduction is the prospective analysis and application of the results of critical appraisal of formal randomized clinical trials. This method of decision making will successfully withstand formal attempts to demonstrate the worthlessness of a proven therapy. Therapy proven by well-done RCTs is what physicians should be doing for their patients, and it is what medical students should integrate into clinical practice for the rest of their professional lives. One note of caution belongs here. It is not possible to have an RCT for every question about medicine. Some diseases are so rare or therapies so dangerous that it is unlikely that a formal large RCT will ever be done to answer that clinical query. For these types of questions, observational studies or less rigorous forms of evidence may need to be applied to patients.

There are three global issues to identify when evaluating an RCT (Table 15.1). These are (1) the ultimate objective of treatment, (2) the nature of the specific treatment, and (3) the treatment target. The ultimate objective of treatment must be defined before the commencement of the trial. While we want therapy to cure and eliminate all traces of disease, more often than not other outcomes will be sought. Therapy can reduce mortality or prevent a treatable death, prevent recurrence, limit structural or functional deterioration, prevent later complications, relieve the current distress of disease including pain in the terminal phase of illness, or deliver reassurance by confidently estimating the prognosis. These are all very different goals and any study should specify which ones are being sought.

After deciding on the specific outcome one wishes to achieve, one must then decide which element of sickness the therapy will most affect. This is not always the disease or the pathophysiologic derangement itself. It may be the illness experience of the patient or how that

pathophysiologic derangement affects the patient through the production of certain signs and symptoms. Finally, it could also be how the illness directly or indirectly affects the patient through disruption of the social, psychological, and economic function of their lives.

Characteristics of RCTs

The majority of RCTs are drug studies or studies of therapy. Often, researchers or drug companies are trying to prove that a new drug is better than drugs that are currently in use for a particular problem. Other researched treatments can be surgical operations, physical or occupational therapy, procedures, or other modalities to modify illness. We will use the example of drug trials for most of this discussion. However, any other medical question can be substituted for the subject of an RCT. The basic rules to apply to critically evaluate RCTs are covered in the following pages.

Hypothesis

The study should contain a hypothesis regarding the use of the drug in the general medical population or the specific population tested. There are two basic types of drug study hypotheses. First, the drug can be tested against placebo, or second, the drug can be tested against another regularly used active drug for the same indication. “Does the drug work better than nothing?” looks at how well the drug performs against a placebo or inert treatment. The placebo effect has been shown to be relatively consistent over many studies and has been approximated to account for up to 35% of the treatment effect. A compelling reason to compare the drug against a placebo would be in situations where there is a question of the efficacy of standard therapies. The use of Complementary and Alternative Medicines (CAM) is an example of testing against placebo and can often be justified since the CAM therapy is expected to be less active than standard medical therapy. Testing against placebo would also be justified if the currently used active drug has

never been rigorously tested against active therapy. Otherwise, the drug being tested should always be compared against an active drug that is in current use for the same indication and is given in the correct dose for the indication being tested.

The other possibility is to ask “Does the drug work against another drug which has been shown to be effective in the treatment of this disease in the past?” Beware of comparisons of drugs being tested against drugs not commonly used in clinical practice, with inadequate dosage, or uncommon routes of administration. These caveats also apply to studies of medical devices, surgical procedures, or other types of therapy. Blinding is difficult in studies of modalities such as procedures and medical devices, and should be done by a non-participating outside evaluation team. Another way to study these modalities is by ‘expert based’ randomization. In this method, various practitioners are selected as the basis of randomization and patients enrolled in the study are randomized to the practitioner rather than the modality.

When ranking evidence, the well-done RCT with a large sample size is the highest level of evidence for populations. A subgroup of RCTs called the n-of-1 trial is stronger evidence in the individual patient and will be discussed later. An RCT can reduce the uncertainty surrounding conflicting evidence obtained from lesser quality studies as illustrated in the following example. Over the past 20 years, there were multiple studies that demonstrated decreased mortality if magnesium was given to patients with acute myocardial infarction (AMI). Most of these studies were fairly small and showed no statistically significant improvement in survival. However, when they were combined in a single systematic review, also called a meta-analysis, there was definite statistical and clinical improvement. Since then, a single large randomized trial called ISIS-4, enrolled thousands of patients with AMI and showed no

beneficial effect attributable to giving magnesium. It is therefore very unlikely that magnesium therapy would benefit AMI patients.

RCTs are the strongest research design capable of proving cause-and-effect relationships. The cause is often the treatment, preventive medicine, or diagnostic test being studied, and the effect is the outcome of the disease being treated, disease prevention by early diagnosis or disease diagnosis by a test. The study design alone does not guarantee a quality study and a poorly designed RCT can give false results. Thus, just like all other studies, critical evaluation of the components is necessary before accepting the results.

The hypothesis is usually found at the end of the introduction. Each study should contain at least one clearly stated, unambiguous hypothesis. One type of hypothesis to be aware of is a single hypothesis attempting to prove multiple cause-and-effect relationships. This cannot be analyzed with a single statistical test and will lead to data dredging. Multiple hypotheses can be analyzed with multivariate analysis and the risks noted in Chapter 14 should be considered when analyzing these studies. The investigation should be a direct test of the hypothesis, although occasionally it is easier and cheaper to test a substitute hypothesis. For example, drug A is studied to determine its effect in reducing cardiovascular mortality, but what is measured is its effect on exercise-stress-test performances. In this case, the exercise-stress-test performance is a surrogate outcome and is not necessarily related to the outcome in which most patients are interested, mortality.

Inclusion and exclusion criteria

Inclusion and exclusion criteria for subjects should be clearly spelled out so that anyone reading the study can replicate the selection of patients. These criteria ought to be sufficiently broad to allow generalization of the study results from the study sample to a large segment of the

population. This concept is also called **external validity** and was discussed in Chapter 8. The source of patients recruited into the study should minimize sampling or referral bias. For instance, patients selected from specialty health care clinics often are more severely ill or have more complications than most patients. They are not typical of all patients with a particular disease so the results of the RCT may not be generalizable to all patients with the disorder. A full list of the reasons for patients' exclusion, the number of patients excluded for each reason, and the methods used to determine exclusion criteria must be defined in the study. Additionally, these reasons should have face validity. Commonly accepted exclusions are patients with rapidly fatal diseases that are unrelated to the disease being studied, those with absolute or relative contraindications to the therapy, and those likely to be lost to follow-up. Beware if there are too many subjects excluded without sound reasons, as this may be a sign of bias.

Randomization

Randomization is the key to the success of the RCT. The main purpose of randomization is to create study groups that are equivalent in every way except for the intervention being studied. Proper randomization means subjects have an equal chance of inclusion into any of the study groups. By making them as equal as possible, the researcher seeks to limit potential confounding variables. If these factors are equally distributed in both groups, bias due to them is minimized.

Some randomization schemes have the potential for bias. The date of admission to hospital, location of bed in hospital (Berkerson's bias), day of birth, and common physical characteristics such as eye color, all may actually be confounding variables and result in unequal qualities of the groups being studied. The first table in most research papers is a comparison of baseline variables of the study and control groups. This documents the adequacy of the randomization process. In addition, statistical tests should be done to show the absence of

statistically significant differences between groups. Remember that the more characteristics looked at, the higher the likelihood that one of them will show differences between groups, just by chance alone. The characteristics listed in this first table should be the most important ones or those most likely to confound the results of the study.

Allocation of patients to the randomization scheme should be concealed. This means that the process of randomization itself is completely blinded. If a researcher knew to which study group the next patient was going to be assigned, it would be possible to switch their group assignment. This can have profound effects on the study results, acting as a form of selection bias. Patients who appeared to be sicker could be assigned to the study group preferred by the researcher, resulting in better or worse results for that group. Current practice requires that the researcher states whether allocation was concealed. If this is not stated, it should be assumed that it was not done and the effect of that bias assessed.

There are two new randomization schemes that merit consideration as methods of solving more and more complex questions of efficacy. The first is to allow all patients requiring a particular therapy to choose whether they will be randomized or be able to freely choose their own therapy. The researchers can then compare the group that chose randomization with the group that chose to self select their therapy. This has an advantage if the outcome of the therapy being studied has strong components of quality of life measures. It answers the question of whether the patient's choice of being in a randomized trial has any effect on the outcome when compared with the possibility of having either of the therapies being studied as a choice. Another method of randomization is using expertise as the point of randomization. In this method, the patients are not randomized, but the therapy is randomized by the provider, with one provider or group of providers using one therapy and another, the comparison therapy. This is a

useful method for studying surgical techniques or complementary and alternative medicine therapies.

Blinding

Blinding prevents confounding variables from affecting the results of a study. If all subjects, treating clinicians, and observers are blinded to the treatment being given during the course of the research, any subjective effects that could lead to biased results are minimized. Blinding prevents observer bias, contamination, and cointervention bias in either group. Lack of blinding can lead to finding an effect where none exists, or vice versa. No matter how honest, researchers may subconsciously tend to find what they want to find. Ideally, tests for adequacy of blinding should be done in any RCT. The simplest test is to ask participants if they knew which therapy they were getting. If there is no difference in the responses between the two groups, the blinding was successful and there is not likely to be any bias in the results due to lack of blinding.

Some types of studies make blinding challenging, although they can be done. Studies of different surgical methods or operations can be done with blinding by using sham operations. This has been successfully performed and in some cases found that standard therapeutic surgical procedures were not particularly beneficial. A recent series of studies showed that when compared to sham arthroscopic surgery for osteoarthritis, actual arthroscopic surgery had no benefit on outcomes such as pain and disability. Similar use of sham with acupuncture showed an equal degree of benefit from real acupuncture and sham acupuncture, with both giving better results than patients treated with no acupuncture. A recent review of studies of acupuncture for low back pain found that there was a dramatic effect of blinding on the outcomes of the studies. The non-blinded studies found acupuncture to be relatively useful for the short-term treatment of

low back pain with a very low NNTB. However, when blinded studies were analyzed, no such effect was found and the results, presented in Table 15.2, were not statistically significant.¹

Description of methods

The methods section should be so detailed that the study could be duplicated by someone uninvolved with the study. The intervention must be well described, including dose, frequency, route, precautions, and monitoring. The intervention also must be reasonable in terms of current practice since if the intervention being tested is being compared to a non-standard therapy, the results will not be generalizable. The availability, practicality, cost, invasiveness, and ease of use of the intervention will also determine the generalizability of the study. In addition, if the intervention requires special monitoring it may be too expensive and difficult to carry out and therefore, impractical in most ordinary situations.

Instruments and measurements should be evaluated using the techniques discussed in Chapter 7. Appropriate outcome measures should be clearly stated, and their measurements should be reproducible and free of bias. Observers should be blinded and should record objective outcomes. If there are subjective outcomes measured in the study, use caution. Subjective outcomes don't automatically invalidate the study and observer blinding can minimize bias from subjective outcomes. Measurements should be made in a manner that assures consistency and maximizes objectivity in the way the results are recorded. For statistical reasons, beware of composite outcomes, subgroup analysis, and post-hoc cut-off points, which can all lead to Type I errors.

¹ E. Ernst & A. R. White. Acupuncture for back pain: a meta-analysis of randomised controlled trials.

Arch. Intern. Med. 1998; 158: 2235–2241.

The study should be clear about the method, frequency, and duration of patient follow-up. All patients who began the study should be accounted for at the end of the study. This is important because patients may leave the study for important reasons such as death, treatment complications, treatment ineffectiveness, or compliance issues, all of which will have implications on the application of the study to a physician's patient population. A study attrition rate of > 20% is a rough guide to the number that may invalidate the final results. However, even a smaller percentage of patient drop-outs may affect the results of a study if not taken into consideration. The results should be analyzed with an intention-to-treat analysis or using a best case/worst case analysis.

Analysis of results

The preferred method of analysis of all subjects when there has been a significant drop-out or crossover rate is to use an intention-to-treat methodology. In this method, all patient outcomes are counted with the group to which the patient was originally assigned even if the patient dropped out or switched groups. This approximates real life where some patients drop out or are non-compliant for various reasons. Patients who dropped out or switched therapies must still be accounted for at the end of the trial since if their fates are unknown, it is impossible to accurately determine their outcomes. Some studies will attempt to use statistical models to estimate the outcomes that those patients should have had if they had completed the study, but the accuracy of this depends on the ability of the model to mimic reality.

A good example of intention-to-treat analysis was in a study of survival after treatment with surgery or radiation for prostate cancer. The group randomized to radical prostatectomy surgery or complete removal of the prostate gland, did much better than the group randomized to either radiation therapy or watchful waiting with no treatment. Some patients who were initially

randomized to the surgery arm of the trial were switched to the radiation or watchful waiting arm of the trial when, during the surgery, it was discovered that they had advanced and inoperable disease. These patients should have been kept in their original surgery group even though their cancerous prostates were not removed. When the study was re-analyzed using an intention-to-treat analysis, the survival in all three groups was identical. Removing those patients biased the original study results since patients with similarly advanced cancer spread were not removed from the other two groups.

Another biased technique involves removing patients from the study. Removing patients after randomization for reasons associated with the outcome is patently biased and grounds to invalidate the study. Leaving them in the analysis as an intention-to-treat is honest and will not inflate the results. However, if the outcomes of patients who left the study are not known, a best case / worst case scenario should be applied and clearly described so that the reader can determine the range of effects applicable to the therapy.

In the best case/worst case analysis, the results are re-analyzed considering that all patients who dropped out or crossed over had the best outcome possible or worst outcome possible. This should be done by adding the drop-outs of the intervention group to the successful patients in the intervention group and at the same time subtracting the dropouts of the comparison group from the successful patients in that group. The opposite process, subtracting drop out patients from the intervention group and adding them to the comparison group, should then be done. This will give a range of possible values of the final effect size. If this range is very large, we say that the results are **sensitive** to small changes that could result from drop-outs or crossovers. If the range is very small, we call the results **robust**, as they are not likely to change drastically because of drop-outs or crossovers.

Compliance with the intervention should be measured and noted. Lack of compliance may influence outcomes since the reason for non-compliance may be directly related to the intervention. High compliance rates in studies may not be duplicated in clinical practice. Other clinically important outcomes that should be measured include adverse effects, direct and indirect costs, invasiveness, and monitoring of an intervention. A blinded and independent observer should measure these outcomes, since if the outcome is not objectively measured, it may limit the usefulness of the therapy. Remember, no adverse effects among n patients could signify as many as $3/n$ adverse events in actual practice.

Results should be interpreted using the techniques discussed in the sections on statistical significance (Chapters 9–12). Look for both statistical and clinical significance. Look at confidence intervals and assess the precision of the results. Remember, narrow CIs are indicative of precise results while wide CIs are imprecise. Determine if any positive results could be due to Type I errors. For negative studies determine the relative likelihood of a Type II error.

Discussion and conclusions

The discussion and conclusions should be based upon the study data and limited to settings and subjects with characteristics similar to the study setting and subjects. Good studies will also list weaknesses of the current research and offer directions for future research in the discussion section. Also, the author should compare the current study to other studies done on the same intervention or with the same disease.

In summary, no study is perfect, all studies have flaws, but not all flaws are fatal. After evaluating a study using the standardized format presented in this chapter, the reader must decide if the merits of a study outweigh the flaws before accepting the conclusions as valid.

Further problems

A study published in *JAMA* in February 1995 reviewed several systematic reviews of clinical trials, and found that if the trials were not blinded or the results were incompletely reported there was a trend of showing better results.² This highlights the need for the reader to be careful in evaluating these types of trials. Always look for complete randomization, total double blinding, and reporting of all potentially important outcomes. An example of this phenomenon can be seen in the systematic review of studies of acupuncture for back pain that was described earlier.

L'Abbé plots are a graphic technique for presenting the results of many individual clinical trials.³ The plot provides a simple visual representation of all the studies of a particular clinical question. It is a way of looking for the presence of bias in the studies done on a single question. The plot shows the proportion of patients in each study who improved taking the control therapy against the proportion who improved taking the active treatment. Each study is represented by one point and the size of the circle around that point is proportional to the sample size of the study. The studies closest to the diagonal show the least effect of therapy, and farther from the diagonal show a greater effect. In addition to getting an idea of the strength of the difference between the two groups, one can also look for the effects of blinding, sample size, or any other factor on the study results. Figure 15.1 shows the results of studies of the effectiveness of acupuncture on short-term improvements in back pain. The studies are divided by blinded vs.

² K. F. Schulz, I. Chalmers, R. J. Hayes & D. G. Altman. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; 273: 408–412.

³ K. A. L'Abbé, A. S. Detsky & K. O'Rourke. Meta analysis in clinical research. *Ann. Intern. Med.* 1987; 107: 224–233.

non-blinded and by size of sample. One can clearly see that the results of the blinded trials were less spectacular than the unblinded ones.

The n-of-1 trial

An n-of-1 trial is done like any other experiment, but with only one patient as a subject. Some have called this the highest level of evidence available. However, it is only useful in the one patient to whom it is applied. It is a useful technique to determine optimal therapy in a single patient when there appears to be no significant advantage of one therapy over another based on reported clinical trials. In order to justify the trial, the effectiveness of therapy must really be in doubt, the treatment should be continued long-term if it is effective, and the patient must be highly motivated to allow the researcher to do an experiment on them. It is helpful if there is a rapid onset of action of the treatment in question and rapid cessation when treatment is discontinued. There should be easily measurable and clinically relevant outcome measures and sensible criteria for stopping the trial.

Additionally, the patient should give informed consent before beginning the trial. The researcher must have a willing pharmacist and pharmacy that can dispense identical, unlabeled active and placebo or comparison medications. Endpoints must be measurable with as much objectivity as possible. Also, the patient should be asked if they knew which of the two treatments they were taking and a statistician should be available to help evaluate the results.⁴

⁴ For more information on the n-of-1 RCT, consult D. L. Sackett, R. B. Haynes, G. H. Guyatt & P.

Tugwell. *Clinical Epidemiology: a Basic Science for Clinical Medicine*. 2nd edn. Boston, MA: Little Brown, 1991, pp. 225–238.

A user's guide to the randomized clinical trial of therapy or prevention

The following is a standardized set of methodological criteria for the critical assessment of a randomized clinical trial article looking for the best therapy which can be used in practice. It is based, with permission, upon the Users' Guides to the Medical Literature published by *JAMA*.⁵ The University of Alberta (www.med.ualberta.ca/ebm) has online worksheets for evaluating articles of therapy that use this guide.

(1) Was the study valid?

- (a) Was the assignment of patients to treatments really randomized?
 - (i) Was similarity between groups documented?
 - (ii) Was prognostic stratification used in allocation?
 - (iii) Was there allocation concealment?
 - (iv) Were both groups of patients similar at the start of the study?
- (b) Were all patients who entered the study accounted for at its conclusion?
 - (i) Was there complete follow-up of all patients?
 - (ii) Were drop-outs, withdrawals, non-compliers, and those who crossed over handled appropriately in the analysis?
- (c) Were the patients, their clinicians and the study personnel including recorders or measurers of outcomes, blind to the assigned treatment?
- (d) Were the baseline factors the same in both groups at the start of the trial?

⁵ G. Guyatt & D. Rennie (eds.). *Users' Guides to the Medical Literature: a Manual for Evidence-Based Practice*. Chicago: AMA, 2002. See also Bibliography.

- (e) Aside from the intervention being tested, were the two groups of patients treated in an identical manner?
 - (i) Was there any contamination?
 - (ii) Were there any cointerventions?
 - (iii) Was the compliance the same in both groups?

(2) **What are the results?**

- (a) How large was the effect size and were both statistical and clinical significance considered? How large is the treatment effect?
 - (i) If statistically significant, was the difference clinically important?
 - (ii) If not statistically significant, was the study big enough to show a clinically important difference if it should occur?
 - (iii) Was appropriate adjustment made for confounding variables?
- (b) How precise are the results? What is the size of the 95% confidence intervals?

(3) **Will the results help me care for my patient?**

- (a) Were the study patients recognizably similar to my own?
 - (i) Are reproducibly defined exclusion criteria stated?
 - (ii) Was the setting primary or tertiary care?
- (b) Were all clinically relevant outcomes reported or at least considered?
 - (i) Was mortality as well as morbidity reported?
 - (ii) Were deaths from all causes reported?
 - (iii) Were quality-of-life assessments conducted?
 - (iv) Was outcome assessment blind?
- (c) Is the therapeutic maneuver feasible in my practice?

- (i) Is it available, affordable, and sensible?
- (ii) Was the maneuver administered in an adequately blinded manner?
- (iii) Was compliance measured?
- (d) Are the benefits worth the costs?
 - (i) Can I identify all the benefits and costs, including non-economic ones?
 - (ii) Were all potential harms considered?

The CONSORT statement

Beginning in 1993, the Consolidated Standards of Reporting Trials Group, known as the CONSORT group began their attempt to standardize and improve the reporting of the process of randomized clinical trials. This was as a result of laxity of reporting of the results of these trials. Currently most medical journals require that the CONSORT requirements be followed in order for a RCT to be published. Look for the CONSORT flow diagram at the start of any RCT and be suspicious that there are serious problems if there is no flow diagram for the study. The CONSORT flow diagram is outlined in Table 15.3.

Ethical issues

Finally, there are always ethical issues that must be considered in the evaluation of any study. Informed consent must be obtained from all subjects. This is a problem in some resuscitation studies, where other forms of consent such as substituted or implied consent may be used. Look for Institutional Review Board (IRB) approval of all studies. If it is not present, it may be an unethical study. It is the responsibility of the journal to publish only ethical studies. Therefore most journals will not publish studies without IRB approval. Decisions about whether or not to use the results of unethical studies are very difficult and beyond the scope of this book. As always, in the end, readers must make their own ethical judgment about the research.

All the major medical journals now require authors to list potential conflicts of interest with their submissions. These are important to let the reader know that there may be a greater potential for bias in these studies. However, there are always potential reasons to suspect bias based upon other issues that may not be so apparent. These include the author's need to "publish or perish," desire to gain fame, and belief in the correctness of a particular hypothesis. A recent study on the use of bone-marrow transplantation in the treatment of stage 3 breast cancers showed a positive effect of this therapy. However, some time after publication, it was discovered that the author had fabricated some of his results, making the therapy look better than it actually was.

All RCTs should be described, before initiation of the research, in a registry of clinical trials. This can be seen on the ClinicalTrials.gov website, a project of the National Institutes of Health in the United States. This is a registry of clinical trials conducted around the world. The site gives information about the purpose of the clinical trial, who may participate, locations, and phone numbers for more details. These details should be adequate for anyone else to duplicate the trial. The purpose of the registry is to get the details of the trial published on-line prior to initiation of the trial itself. This way, the researchers cannot spin the results to look better by reporting different outcomes than were originally specified or by using different methods than originally planned. Most journals will no longer publish trials that are not registered in this or a similar international registry.

The question of placebo controls is one ethical issue which is constantly being discussed. Since there are therapies for almost all diseases, is it ever ethical to have a placebo control group? This is still a contentious area with strong opinions on both sides. One test for the suitability of placebo use is clinical equipoise. This occurs when the clinician is unsure about the

suitability of a therapy and there is no other therapy that works reasonably well to treat the condition, placebo therapy can be used. Both the researcher and the patient must be similarly inclined to choose either the experimental or a standard therapy. If this is not true, placebo ought not to be used.

Fig. 15.1 Effect of blinding and sample size on results in trials of acupuncture for low back pain. From E. Ernst & A. R. White. *Arch. Intern. Med.* 1998; 158: 2235–2241.

Table 15.1. Schema for randomized clinical trials

Ultimate objective	Specific treatment	Target disorder
cure	drug therapy	disease
reduce mortality	Surgery	illness
prevent recurrence	other therapies	predicament
limit deterioration	Nutrition	
prevention	psychological support	
relieve distress		
deliver reassurance		
allow the patient to die comfortably		

Table 15.2. Effects of acupuncture on short-term outcomes in back pain

Type	Number of trials	Improved with Acupuncture (%)	Improved with control (%)	Relative Benefit (95 % CI)	NNT (95% CI)
Blinded	4	73/127 (57)	61/123 (50)	1.2 (0.9 to 1.5)	13 (5 to no benefit)

Non-blinded	5	78/117 (67)	33/87 (38)	1.8 (1.3 to 2.4)	3.5 (2.4 to 6.5)
-------------	---	-------------	------------	------------------	------------------

Table 15.3: Template for the CONSORT format of an RCT showing the flow of participants through each stage of the study.

1. Assessed for eligibility (n = ...)
2. Enrollment: Excluded (n = ...) Not meeting inclusion criteria (n = ...), Refused to participate (n = ...), Other reasons (n = ...)
3. Randomized (n = ...)
4. Allocation: Allocated to intervention (n = ...), Received allocated intervention (n = ...), Did not receive allocated intervention (n = ...) (give reasons)
5. Follow-up: Lost to follow up (n = ...) (give reasons), Discontinued intervention (n = ...) (give reasons)
6. Analysis: Analyzed (n = ...), Excluded from analysis (n = ...) (give reasons)