CHAPTER 1

Introduction

A very large number of clinical trials with human subjects have been conducted in a wide variety of contexts. Many of these have been concerned, for example, with improving (in some way) the management of patients with disease and others the prevention of the disease or condition in the first place. The essence of a clinical trial is the comparison of a standard strategy with an alternative (perhaps novel) intervention. The aim of this chapter is to illustrate some of the wide variety of clinical trials that have been conducted and to highlight some key features of their design, conduct and analysis.

1.1 Introduction

The aim of this book is to introduce those who are to become involved with randomised clinical trials to the wide range of challenges that are faced by those who conduct such trials. Thus, our intended readership is expected to range from healthcare professionals of all disciplines who are concerned with patient care to those more involved with the non-clinical aspects such as the statistical design, data processing and subsequent analysis of the results. We assume no prior knowledge of clinical trial processes, and we have attempted to explain the more statistical sections in as non-technical a way as possible. In a first reading of this book, these sections could be omitted. Throughout the book, we stress the collaborative nature of clinical trials activity and would hope that readers would consult their more experienced colleagues on aspects of our coverage.

The business of clinical trials is an ongoing process, and as we write, trials are currently being designed (particularly with respect to the coronavirus), opened, conducted, closed, analysed, reported, results filtered into current practice and the next planned. To describe the key features of this process, it is difficult to know where to start as each stage interacts with each of the others to some extent. For example, in designing a trial the investigators need to be mindful of the eventual analysis to be undertaken as this governs (but it is only one aspect of) how large a trial should be launched. Some of the steps are intellectually challenging, for example, defining the key therapeutic question, whilst

Randomised Clinical Trials: Design, Practice and Reporting, Second Edition. David Machin, Peter M. Fayers, and Bee Choo Tai.

others may perhaps appear more mundane, such as defining the data forms or the data entry procedures but all steps (whether large or small: major or minor) underpin the eventual successful outcome – the influence on clinical practice once the trial results are available. For many of these aspects of the process, whole books have been written. We can only provide an introduction to these.

Numerous terms including 'clinical trial' itself need to be introduced. As a consequence, we have included a Glossary of Terms, which is mainly extracted from Day (2007) *Dictionary of Clinical Trials*. Thus, the Glossary defines: **clinical trial**: any systematic study of the effects of a treatment in human subjects. These definitions may not be exhaustive in the sense, that 'treatment' used here may be substituted by, for example, 'intervention' depending on the specific context of the clinical trial under consideration.

Clinical trials require a multidisciplinary approach in which all partners play a key role at some stage of the trial process. Furthermore, 'Evidence-Based Medicine' (EBM) requires that it is important to consider critically *all* the available evidence about whether, for example, a treatment works, before recommending it for clinical practice. In this respect, it is therefore vital that one can clearly see that a proposed trial addresses a key question which will have a clinically meaningful outcome, is well designed, conducted and reported, and the results are persuasive enough to change clinical practice if appropriate.

Despite perhaps not having a professional interest in the science of clinical trials, everyone has a vested interest in them as potential patients requiring care. How many of us have never been to see a doctor, had a hospital admission or taken medication? All of us may be, have been, or certainly will be, recipients of clinical trial results whether during prebirth, at birth or in childhood for vaccination and minor illness, as an adult for fertility, sports injuries, minor and major non-life-threatening or life-threatening illnesses, and in old age for care related to our mental or physical needs.

1.2 Some completed trials

As we have indicated, there are countless ongoing trials and many have been successfully conducted and reported. To give some indication of the range and diversity of application, we describe a selection of clinical trials that have been conducted. Their designs include some features that we also draw upon as examples in later chapters.

Example 1.1 Small parallel two-group design – gastrointestinal function

Lobo, Bostock, Neal, *et al.* (2002) describe a randomised trial in which 20 patients with colonic cancer either received postoperative intravenous fluids in accordance with current hospital standard practice (*S*) or according to a restricted intake regimen (*R*). A primary endpoint measure in each patient was the solid-phase gastric

Example 1.1 (*Continued*)

emptying time on the fourth postoperative day. The observed difference between the median emptying times was shorter with *R* by 56 minutes with 95% confidence interval (CI) from 12 to 132 minutes. The trial also included preoperative and postoperative (days 0, 1, 3 and 5) measures of the concentrations of serum albumin, haemoglobin and blood urea in a repeated measures design.

Key features include the following:

- *Design:* Randomised comparison of a standard and test, single-centre participation, unblinded assessment,
- *Endpoint:* Gastric emptying time,
- *Size:* 21 patients following colonic resection,
- *Analysis:* Mann–Whitney *U*-test[1] for comparing two medians,
- *Conclusion:* The restricted intake group had shorter delays in returning to gastrointestinal function.

[1]This can also be referred to as the Wilcoxon rank-sum test.

Example 1.2 Parallel two-group design – hepatitis B

Levie, Gjorup, Skinhøj and Stoffel (2002) compared a 2-dose regimen of recombinant hepatitis B vaccine including the immune stimulant AS04 with the standard 3-dose regimen of HbsAg in healthy adults. The rationale behind testing a 2-dose regimen was that fewer injections would improve compliance.

Key features include the following:

- *Design*: Two centres, open-label randomised two-group comparison,
- *Endpoint*: Seroprotection rate,
- *Size*: 340 healthy adults aged between 15 and 40 years,
- *Analysis*: Fisher's exact test,
- *Conclusion*: The 2-dose regimen compared favourably with the standard.

Example 1.3 Unstructured three-group design – newly diagnosed type 2 diabetes

The randomised trial of Weng, Li, Xu, *et al.* (2008) compared, in newly diagnosed patients with type 2 diabetes, three treatments: multiple daily insulin injections (MDI), continuous subcutaneous insulin infusion (CSII) and oral hypoglycaemic agent (OHA).

Key features include the following:

- *Design*: Nine centres, randomised three-group comparison,
- *Endpoint*: Time of glycaemic remission,
- *Size*: 410 newly diagnosed patients with type 2 diabetes,
- *Analysis*: Cox proportional-hazards regression model,
- *Conclusion*: Early intensive therapy has favourable outcomes on recovery and maintenance of β-cell function and protracted glycaemic remission compared to OHA.

Example 1.4 Small dose–response design – pain prevention following hand surgery

Stevinson, Devaraj, Fountain-Barber, *et al.* (2003) conducted a randomised double-blind, placebo-controlled trial to compare placebo with homoeopathic arnica 6C and arnica 30C to determine the degree of pain prevention in patients with carpel tunnel syndrome undergoing elective surgery for their condition. Pain was assessed postoperatively with the short-form McGill Pain Questionnaire (SF-MPQ) at four days. A total of 64 patients were randomised to the three groups resulting in median scores of 16.0 (range 0–69), 10.5 (0–76) and 15.0 (0–82) for the respective groups. From these results, the authors suggest that homoeopathic arnica has no advantage over placebo in reducing levels of postoperative pain.

Key features include the following:

- *Design*: Single-centre, randomised double-blind, placebo-controlled, three-group dose response,
- *Endpoint*: Pain using the SF-MPQ,
- *Size*: 64 patients undergoing hand surgery for carpal tunnel syndrome,
- *Analysis*: Kruskal–Wallis test,
- *Conclusion*: Irrespective of dose homoeopathic arnica has no advantage over placebo.

Example 1.5 Large dose–response design – HER2-positive breast cancer

Smith, Procter, Gelber, *et al.* (2007) showed that 1 year of treatment with Trastuzumab (*T*) after adjuvant therapy in HER2-positive patients with breast cancer was superior to Observation (*O*) alone. They reported a hazard ratio, $HR = 0.66$ (95% CI 0.47 to 0.91, p-value = 0.0115) for overall survival in favour of adjuvant treatment. This comparison was from two arms of a three-arm large multicentre international randomised trial comprising 1698 patients randomised to *O*, 1703 to *T* for 1 year (*T*1) and 1701 to *T* for 2 years (*T*2): a total of 5102 patients.

Key features include the following:

- *Design*: Randomised, multicentre, observation versus active treatment,
- *Size*: Part of a large trial of 5102 women with HER2-positive breast cancer,
- *Endpoint*: Overall survival,
- *Analysis*: Comparison in 3404 women from the *O* and *T*1 groups using survival curves,
- *Conclusion*: Treatment with *T*1 after adjuvant chemotherapy has a significant overall survival benefit.

Example 1.6 Non-inferiority trial – uncomplicated falciparum malaria

Zongo, Dorsey, Rouamba, *et al.* (2007) conducted a randomised non-inferiority trial to test the hypothesis that the risk of recurrent parasitaemia was not significantly worse with artemether–lumefantrine (*AL*) than with amodiaquine plus sulfadoxine–pyrimethamine (*AQ* + *SP*). A total of 826 patients were screened of which 548 were found to have uncomplicated malaria and were randomised (273 to *AQ* + *SP* and 275 to *AL*). A primary endpoint was the risk of treatment failure within 28 days of randomisation. The authors concluded that *AQ* + *SP*, with a recurrent malaria rate of 1.7% (4/233), was more effective than *AL*, with a rate of 10.2% (25/245) and representing a difference of 8.5% (95% CI 4.3–12.6%). These results suggest that the hypothesis of 'non-inferiority' should not be accepted as the CI included the non-inferiority limit of 3% set by the investigators.

Key features include the following:

- *Design*: Multicentre, two-group, non-inferiority trial,
- *Endpoint*: Time to recurrent malaria,
- *Size*: Large – 548 patients with uncomplicated falciparum malaria,
- *Analysis*: Comparison of Kaplan–Meier survival curves,
- *Conclusion*: *AL* was less effective than (inferior to) *AQ* + *SP*.

Example 1.7 Repeated measures – atopic eczema

Meggitt, Gray and Reynolds (2006) randomised 63 patients with moderate-to-severe atopic eczema to receive either Azathioprine or Placebo in a double-blind formulation to ascertain the relative reduction in disease activity determined by the six-area six-sign atopic dermatitis (SASSAD) score between the groups. One patient in each group subsequently withdrew from the trial before treatment was initiated. The investigators reported a 5.4 unit advantage with Azathioprine. In this trial, patients were randomised using a minimisation procedure, in the ratio of 2 to 1 in favour of Azathioprine in order to.

> … encourage recruitment, to reduce the numbers receiving pharmacologically inactive systemic treatment, and to increase the likelihood of identifying infrequent adverse events.

Key features include the following:

- *Design*: Single-centre, randomised double-blind, placebo-controlled, randomised 2 : 1 allocation ratio using minimisation,
- *Endpoint*: SASSAD,
- *Size:* 63 patients with moderate-to-severe atopic eczema,
- *Analysis:* Comparison of mean group regression slopes over a 12-week period,
- *Conclusion*: Azathioprine produces a clinically relevant improvement.

Example 1.8 Cross-over trial – known or suspected hypertension

Kerley, Dolan, James and Cormican (2018) describe a randomised placebo (*P*) controlled, two-period cross-over trial of dietary nitrate (*N*) in 20 patients with known or suspected hypertension. The *P* and *N* interventions were delivered in beetroot juice in a nitrogen-depleted or nitrogen-enriched form, respectively. Thirteen of the individuals were randomised to receive the sequence *NP*, that is *N* in Period I of the trial followed by *P* in Period II, and the other seven were allocated the sequence *PN*. Amongst the many endpoints, plasma nitrate and ambulatory blood pressure were recorded prior to randomisation, then 7 days later following the start of treatment in Period I and a further 7 days following Period II. The authors concluded:

> Our results support … an anti-hypertensive effect of dietary nitrate … .

Key features include the following:

- *Design*: Single-centre, randomised placebo-controlled, two-period cross-over trial,

Example 1.8 (*Continued*)

- *Size*: 20 patients with known or suspected hypertension, unequal numbers assigned to the sequences,
- *Washout*: None included,
- *Endpoints*: Plasma nitrate and ambulatory blood pressure,
- *Analysis*: Complex methodology described. All statistical tests were conducted at the two-sided 0.05 significance level,
- *Conclusion*: Nitrogen-enriched has an anti-hypertensive effect.

Example 1.9 Paired design – glaucoma

Glaucoma Laser Trial Research Group (1995) recruited 271 subjects with newly diagnosed primary open-angle glaucoma, and one eye of each patient was randomly assigned as initial treatment by argon laser (*L*) trabeculoplasty followed by Stepped (*S*) medication (*LS*). The other eye then received the treatments in reverse order, *SL*. They reported on the 261 eyes and found that measures of visual field status for eyes treated by the sequence *LS* were slightly better than those treated by *SL*. The authors' state:

> Statistical significance was attained for only some of the differences, and the clinical implications of such small differences are not known.

Key features include the following:

- *Design*: Multicentre, paired design, compares alternative schedules for administering two procedures – the schedule was randomised to one eye with the other eye receiving the alternative,
- *Endpoint*: Visual field status,
- *Size*: 271 patients with primary open-angle glaucoma,
- *Analysis*: Comparison of means at particular time points following initiation of treatment using the paired *t*-test,
- *Conclusion*: Eyes treated with laser trabeculoplasty first were slightly better than those eyes treated with topical medication first.

Example 1.10 Split-mouth design – implants for edentulous sites

Pozzi, Agliardi, Tallarico and Barlattani (2012) conducted a trial in 34 partially edentate patients who required at least two single implant-supported crowns. A split-mouth design was used in which one of two different prosthetic interfaces and configurations: internal conical connection with back-tapered collar and platform shifting (*CC*) or external-hexagon implants with flat-to-flat implant-abutment interface (*EH*), were randomly allocated at each edentulous site. From a total of 88 implants included in the trial, the authors concluded that both implants performed similarly in terms of failure rates.

Key features include the following:

- *Design*: Single-centre, split-mouth, random allocation,
- *Endpoint*: Failure rates and marginal bone loss,
- *Size*: 34 patients with 88 edentulous sites,
- *Analysis*: Comparing implants using paired *t*-tests at several intervals postrandomisation,
- *Conclusion*: Lower marginal bone loss with *CC* when compared to *EH*.

Example 1.11 Cluster trial – hip protectors for the elderly

Meyer, Warnke, Bender and Mülhauser (2003) conducted a trial involving 942 residents from 49 nursing homes. In this cluster trial design, the nursing homes contain 'clusters' of residents and the homes (not the individual residents) were randomised, with 25 homes, comprising a total of 459 residents, assigned to the intervention group and 24, with 483 residents, to the usual care (control) group. The intervention comprised a single education session for nursing staff, who then educated residents, and the provision of three hip protectors per resident. The control clusters administered usual care optimised by brief information to nursing staff about hip protectors and the provision of two hip protectors per cluster for demonstration purposes. The main outcome measure was the incidence of hip fractures. There were 21 hip fractures in 21 (4.6%) residents in the intervention group and 42 in 39 (8.1%) residents in the control group – a difference of 3.5% (95% CI 0.3–7.3%, p-value = 0.072). The authors concluded:

> The introduction of a structured education programme and the provision of free hip protectors in nursing homes may reduce the number of hip fractures.

Example 1.11 (*Continued*)

Key features include the following:

- *Design*: Multicluster, randomised,
- *Size*: 49 nursing homes comprising 942 residents with high risk of falling,
- *Endpoint*: Hip fractures,
- *Analysis*: Chi-squared test adjusted for cluster randomisation,
- *Conclusion*: A structured education programme and the provision of hip protectors may reduce the number of hip fractures.

Example 1.12 Single arm – discrete de novo lesions in a coronary artery

Erbel, Di Mario, Bartunek, *et al.* (2007) describe a non-randomised multicentre trial involving 8 centres in which 63 patients were enrolled with single de novo lesions in a native coronary artery. In these patients, a total of 71 biodegradable magnesium stents were successfully implanted. The (composite) primary endpoint was the rate of major adverse cardiac events (MACE) defined as any one of: cardiac death, Q-wave myocardial infarction or target lesion revascularisation at 4 months poststent implant. This was to be compared with an anticipated rate of 30%. They reported a rate of MACE at 4 months of 15/63 (23.8%); all of which were attributed to target lesion revascularisation (there were no deaths or Q-wave myocardial infarctions) and concluded:

> … biodegradable magnesium stents can achieve an immediate angiographic result similar to … other metal stents … .

Nevertheless, the authors also commented in their Discussion:

> The absence of randomisation precludes direct comparison with other techniques of percutaneous revascularisation.

Key features include the following:

- *Design*: No comparison group hence non-randomised, multicentre,
- *Size*: 71 stents in 63 patients,
- *Endpoint*: Composite endpoint – MACE,
- *Analysis*: Proportion experiencing MACE with 95% confidence interval,
- *Conclusion*: Bioabsorbable stents can achieve an immediate angiographic result similar to other metal stents and can be safely degraded.

The above examples of successfully completed clinical trials illustrate a wide range of topics investigated. These include patients with disease (breast cancer, colon cancer, eczema, glaucoma, malaria and diabetes mellitus), those requiring coronary artery stents or hand surgery, elderly residents of nursing homes, patients aged 25 years or more requiring at least two implant-supported crowns for dental caries, healthy individuals and those requiring vaccinations. Although not included here, trials are conducted, for example, to evaluate different diagnostic procedures, different bed mattresses to reduce the incidence of bed sores, different dressings for wounds of all types and fertility regulation options for male and females of reproductive potential.

These trials are often termed Phase III trials in contrast with Phase I and Phase II trials which are concerned with early stages of the (often pharmaceutical) development process. Although the trials differ in aspects of their design, the majority have the general structure of a two (or more) group parallel design in which eligible patients are assigned to receive the alternative options (often treatments but more generally termed interventions) and then at some later time assessed in a way which will be indicative of (successful) outcome. The outcomes measured in these trials include the following: survival time, gastric emptying time, reduction in disease activity, visual field status, recurrent parasitaemia, major adverse cardiac events, pain, the number of hip fractures, systolic blood pressure and standard criteria used to assess dental restorations. In the trial of homoeopathic arnica for pain relief following hand surgery, assessment was made in a double-blind or double-masked manner in which neither the patient nor the assessor was aware of the specific treatment option actually received.

The methods used for the allocation to the options included simple randomisation of equal numbers per group, a 2 to 1 allocation; a minimisation procedure taking into account patient characteristics, randomisation to nursing homes (clusters) rather than to individual residents. For the split-mouth design used for the comparing dental implants the authors' state:

> For randomization of the implant type, a pregenerated random sequence was created … . Opaque envelopes were sealed according to pregenerated list. An independent judge prepared all envelopes. … an assistant indicated which implant had to be placed first following the indications contained in the sequentially number envelope.

The non-random allocation to a single-arm study using a new bioabsorbable stent for coronary scaffolding might now be regarded as a feasibility study although the trial results were compared to that from historical data.

The trials ranged in size from 20 patients with colonic cancer to 5102 women with HER2-positive breast cancer. One trial involved 522 eyes from 271 subjects another 88 single implant-supported crowns teeth in 34 partially edentate patients. Although not fully detailed in the above summaries, methods of statistical analysis ranged from a simple comparison of two proportions to relatively complex methods using techniques for survival time outcomes.

In general, trials are designed to establish a difference between the (therapeutic) options under test and were one to exist. Consequently, they are sometimes termed *superiority* trials. However, in certain circumstances, as in the trial for the treatment of

uncomplicated falciparum malaria, the research team were looking for *non-inferiority* implying that the two treatment strategies of $AQ + SP$ and AL would give very similar risks of failure. In the event, the trial suggested that AL was (unacceptably) less effective than $AQ + SP$ implying that non-inferiority was not established. Such designs usually imply that a satisfactory outcome is that the test treatment does not perform worse than the standard to an extent *predefined* by the investigating team. Thus, use of a non-inferiority design often implies that, although some therapeutic loss may be conceded on the main outcome variable, other factors favouring the new therapy will have some features (*gain*) to offset this. For example, if the new compound was a little less effective (not equal to) but had a better toxicity profile, then this might be sufficient to prefer it for clinical practice.

1.3 Choice of design

1.3.1 Biological variability

Measurements made on human subjects rarely give exactly the same results from one occasion to the next. Even in adults, our height varies a little during the course of the day. If one measures the blood sugar levels of an individual on one particular day and then again the following day, under exactly the same conditions, greater variation in this than that of height would be expected. Hence were such an individual to be assessed and then receive an intervention (perhaps to lower blood sugar levels) any lowering recorded at the next assessment cannot necessarily be ascribed to the intervention itself. The levels of inherent variability may be very high so that, perhaps in the circumstances where a subject has an illness, the oscillations in these may disguise, at least in the early stages of treatment, the beneficial effect of the treatment given to improve the condition.

Example 1.13 Patient-to-patient variability – atopic eczema

The considerable between patient variability in the trial of Example 1.7 is illustrated in Figure 1.1. In the 41 patients receiving Azathioprine, the reduction in disease activity (SASSAD) ranged from −10 to 32. There is considerable overlap of these values with those from the 20 patients receiving Placebo whose values range from −12 to 20. This figure clearly illustrates that, although there is considerable variation, the majority of patients in both groups improve. Further, the corresponding reduction in percentage body area affected with Azathioprine was reported to range from approximately −15 to 85% and for placebo approximately −20 to 45%. Nevertheless, even with the majority of patients improving in both groups, the trial of Meggitt, Gray and Reynolds (2006) indicated a better outcome, on average, for those receiving Azathioprine.

Example 1.13 (*Continued*)

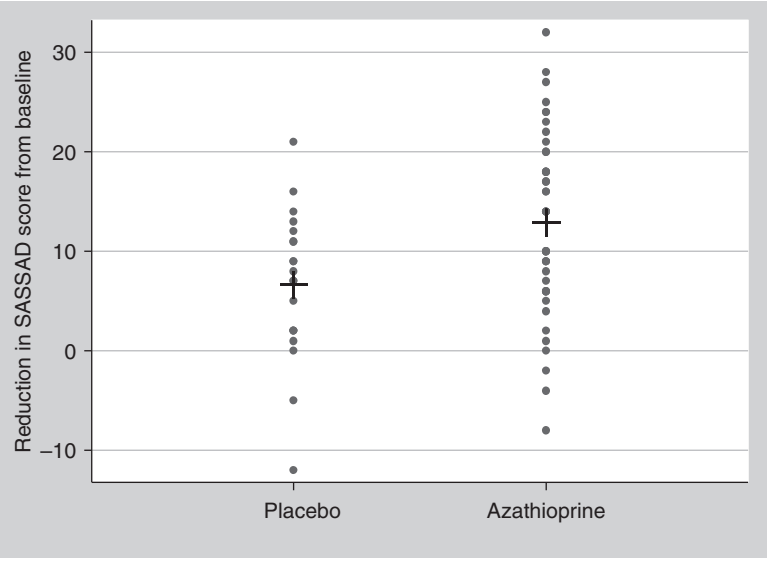

Figure 1.1 Individual patient reductions in disease activity (SASSAD) for the Azathioprine and Placebo treatment groups with the corresponding means indicated. Source: Data from Meggitt, Gray and Reynolds (2006).

With such variability, it follows that, in any comparison made in a biomedical context, differences between subjects or groups of subjects frequently occur. These differences may be due to real effects, random variation or both. It is the job of the experimenter to decide how this variation should be taken note of in the design of the ensuing trial. The purpose being that, once at the analysis stage, the variation can be partitioned suitably into that due to any real effect of the interventions on the difference between groups and that from the random or chance component.

1.3.2 Randomisation

Ronald A Fisher (1890–1962) in laying the foundations of good experimental design, although in an agricultural and biological context, advocated the use of randomisation in allocating experimental treatments. Thus, for example, in agricultural trials various plots in a field are randomly assigned to the different experimental interventions. The argument for randomisation is that it will prevent systematic differences between the allocated plots receiving the different interventions, whether or not these can be identified by the investigator concerned, before the experimental treatment is applied.

Then, once the experimental treatments are applied and the outcome observed, the randomisation enables any differences between treatments to be estimated objectively and without bias. In these and many other contexts, randomisation has long been a keystone to good experimental design.

The need for random allocation extends to all experimental situations including those concerned with patients as opposed to agricultural plots of land. The difficulty arises because clinical trials (more emotive than experiments) do indeed concern human beings who cannot be regarded as experimental units and so should not be allocated the interventions without their consent. The consent process clearly complicates the allocation process and, at least in the past, has been used as a reason to resist the idea of randomisation of patients to treatment. Unfortunately, the other options, perhaps a comparison of patients receiving a 'new' treatment with those from the past receiving the 'old', are flawed in the sense that any observed differences (or lack thereof) may not reflect the true situation. Thus, in the context of controlled clinical trials, Pocock (1983) concluded, many years ago and some 30 years after the first randomised trials were conducted, that:

> The proper use of randomization guarantees that there is no bias in the selection of patients for the different treatments and so helps considerably to reduce the risk of differences in experimental environment. Randomized allocation is not difficult to implement and enables trial conclusions to be more believable than other forms of treatment allocation.

As a consequence, we are focussing on randomised controlled trials and not giving much attention to less scientifically rigorous options.

1.3.3 Design hierarchy

The final choice of design for a clinical trial will depend on many factors, key amongst these are clearly the specific research question posed, the practicality of recruiting patients to such a design and the resources necessary to support the trial conduct. We shall discuss these and other issues pertinent to the design choice in later chapters. Nevertheless, we can catalogue the main types of design options available and these are listed in Figure 1.2. This gives a relative weight to the evidence obtained from these different types of clinical trial. All other things being equal, the design that maximises the weight of the resulting evidence should be chosen. For expository purposes, we assume that a comparison of a new test treatment with the current standard for the specific condition in question is being made.

1.3.3.1 Randomisation

The design that provides the strongest type of evidence is the *double-blind* (*or double-masked*) *randomised controlled trial* (RCT). In this, the patients are allocated to treatment at random and this ensures that *in the long run* patients, before treatment commences, will be comparable in the test and standard groups. Clearly, if the important

Evidence level	Type of trial
Strongest	Double-blind randomised controlled trial (RCT)
	Single-blind RCT
	Non-blinded (open) RCT
	Non-randomised prospective trial
	Non-randomised retrospective trial
	Before-and-after design (historical control)
Weakest	Case-series

Figure 1.2 The relative strength of evidence obtained from alternative designs for comparative clinical trials

prognostic factors that influence outcome were known, one could match the patients in the standard and test groups in some way. However, the advantage of randomisation is that it balances for *unknown* and the *known* prognostic factors and this could not be achieved by matching. Thus, the reason for the attraction of the randomised trial is that it is the *only* design that can give an absolute certainty that there is no bias in favour of one group compared to another at the start of the trial. Indeed, in Example 1.12, Erbel, Di Mario, Bartunek, *et al.* (2007), who essentially conducted a single-arm prospective case study, admitted that failure to conduct a randomised comparison compromised their ability to draw definitive conclusions concerning the stent on test.

1.3.3.2 Blinding or masking

For the simple situation in which the attending clinician is also the assessor of the outcome, the trial should ideally be double-blind (or double-masked). This means that neither the patient nor the attending clinician will know the actual treatment allocated. Having no knowledge of which treatment has been taken, neither the patient nor the clinician can be influenced at the assessment stage by such knowledge. In this way, an unprejudiced evaluation of the patient response is obtained. Thus Meggitt, Gray and Reynolds (2006) used double-blind formulations of Azathioprine or Placebo so that neither the patients with moderate-to-severe eczema, nor their attending clinical team, were aware of who received which treatment. Although they did not give details, the blinding is best broken only at the analysis stage once all the data had been collated.

Despite the inherent advantage of this double-blind design, most clinical trials cannot be conducted in this way as, for example, a means has to be found for delivering the treatment options in an identical way. This may be a possibility if the standard and test are available in tablet form of identical colour, shape, texture, smell and taste. If such

'identity' cannot be achieved, then a single-blind design may ensue. In such a design, one of the patient or the clinical assessor has knowledge of the treatment being given but the other does not. In trials with patient survival time as the endpoint, double-blind usually means that both the patient and the treating physician and other staff are blinded. However, assessment is objective (death) and the blinding irrelevant by this stage.

Finally, and this is possibly the majority situation, there will circumstances in which neither the patient nor the assessor can be blind to the treatments actually received. Such designs are referred to as 'open' or 'open-label' trials.

1.3.3.3 Non-randomised designs

In certain circumstances, when a new treatment has been proposed for evaluation, all patients are recruited prospectively but allocation to treatment is not made at random. In such cases, the comparisons may well be biased and hence are unreliable. The bias arises because the clinical team choose which patients receive which intervention and in so doing may favour (even subconsciously) giving one treatment to certain patient types and not to others. In addition, the requirement that all patients should be suitable for all options may not be fulfilled – in that if it is known that a certain option is to be given to a particular subject then one may not so rigorously check if the other options are equally appropriate. Similar problems arise if investigators have recruited patients into a single-arm study, and the results from these patients are then compared with information on similar patients having (usually in the past) received a relevant standard therapy for the condition in question. However, such historical comparisons are likely to be biased also and to an unknown extent so again it will not be reasonable to ascribe the difference (if any) observed entirely to the treatments themselves. Of course, in either case, there will be situations when one of these designs is the only option available. In such cases, a detailed justification for not using the 'gold standard' of the randomised controlled trial is required.

Understandably, in this era of EBM, information from non-randomised comparative studies is categorised as providing weaker evidence than that from randomised trials.

The before-and-after design is one in which, for example, patients are treated with the *Standard* option for a specified period and then, at some fixed point in time, subsequent patients receive the *Test* treatment. This is the type of design used by Erbel, Di Mario, Bartunek, *et al.* (2007) to evaluate a bioabsorbable stent for coronary scaffolding. In such examples, the information for the *Standard* is retrospective in nature and is often obtained from clinical records only and so was not initially collected for trial purposes. If this is the case, the before-and-after design is likely to be further compromised as, for example, in the 'before' period, the patient selection criterion, clinical assessments and data recorded may not meet the standards required of the 'after' component. Such differences are likely to influence the before-and-after comparison in unforeseen and unknown ways.

Example 1.14 Glioblastoma in the elderly – non-randomised design

Brandes, Vastola, Basso, *et al.* (2003) describe a study comparing radiotherapy alone (Group A), radiotherapy and the combination of procarbazine, lomustine and vincristine (Group B) and radiotherapy with temozolomide (Group C) in 79 elderly patients with glioblastoma. The authors' state:

> The first group (Group A) was enrolled in the period from March 1993 to August 1995 … . The second group (Group B) was enrolled from September 1995 to September 1997 … . The third group (Group C) was enrolled from September 1997 to August 2000 and … .

The authors conclude:

> Overall survival was better in Group C compared with Group A (14.9 months v 11.2 months, P = 0.002), but there was no statistical differences found between Groups A and B or between Groups B and C.

However, since patients have not been randomised to groups, one cannot be sure that the differences (and lack of differences) truly reflect the relative efficacy of the three treatments concerned. This type of design should be avoided if at all possible.

1.3.3.4 Case series

A case series consists of a study in which the experience of an investigator treating a series of patients with a particular approach reports on their outcome. This may be the only 'design' option available in rare or unusual circumstances but is unlikely to provide clear evidence of efficacy. There are many criticisms of this design. Generally one may not know how the patients have been selected; the clinical team may have an eye for selecting those patients to be given the treatment who are likely to recover in any event: without further evidence of the natural history of the disease, we do not know whether the patients may have recovered spontaneously, without intervention: we do not know whether their approach to treatment is better than any alternatives.

1.4 Practical constraints

Control of the 'experiment' is clearly a desirable feature – perhaps easy to attain in the physics laboratory where experimental conditions are tightly controlled but not so easy with living material particularly if they are human. A good trial should answer the questions posed as efficiently as possible. In broad terms, this implies recruiting as few subjects as is reasonably possible for a reliable answer to be obtained.

Although good science may lead to an optimal choice of design, the exigencies of real life may cause these ideals to be modified. All the same, one can still keep in mind the hierarchy in the choice of designs of Figure 1.2, but where to enter this hierarchy will depend on circumstance. Thus, the investigators do not aim for the best design, but only the best realisable design in their context.

Technical (statistical) aspects of experimental design can be used in a whole variety of settings. Nevertheless, there are specific problems associated with implementing these designs in practice in the field of clinical trials. It is clear that trials cannot be conducted without human subjects (often patients); nevertheless, the constraints this imposes are not inconsiderable. Figure 1.3 illustrates some aspects that need to be considered when conducting such trials.

As we have indicated, the requirements for human studies are usually more stringent than in other research areas. For example, safety, in terms of the welfare of the experimental units involved, is of overriding concern in clinical trials, possibly of less relevance in animal studies and of no relevance to laboratory studies. In some sense, the laboratory provides, at least in theory, the greatest rigour in terms of the experimental design, and studies in human subjects should be designed (whenever possible) to be as close to these standards as possible. However, no consent procedures from the experimental units nor from animals, if they are involved, are required, whereas this is a very important consideration in all human experimentation even in a clinical trial with therapeutic intent.

Constraints may also apply to the choice of interventions to compare. For example, in certain therapeutic trials there may be little chance that a placebo option will bring any benefit (although this is certainly not the case in all circumstances) so comparisons

Design feature	
Method of assessments	If invasive – may not be acceptable.
Treatment or Intervention	Implicit that treatment should do some good – thus an innocuous or placebo treatment may not be acceptable.
Subject safety issues	Overriding principle is the safety of the subjects
Protocol Review	Scientific and ethical
Consent	Fully informed consent mandatory
Recruitment	Usually, subjects recruited one-by-one over calendar time
Time scale	May be relatively long – rarely weeks, seldom months, quite often years
Trial size	Not too large or too small
Patient losses	Subjects may refuse to continue in the trial at any stage
Observations	Usually, subjects assessed one-by-one over calendar time
Design changes	Almost certainly require new ethical approval
Data protection	Confidentiality and often National Guidelines for storage and transfer.
Reporting	CONSORT for Phase III trials (Moher, Hopewell, Schultz, *et al*, 2010)

Figure 1.3 Special considerations for clinical trials in human subjects

may have to be made between two allegedly 'active' approaches despite little direct evidence that either of them will bring benefit. However, if, at the end of such a trial, a difference between treatments is demonstrated then activity for the better of the two is established so in one sense comparison with a placebo was not necessary. In contrast should the two treatments appear not to differ in their effectiveness then no conclusions can be drawn since one does not know whether both are equally beneficial or whether both are equally ineffective as compared to *Placebo*. Thus, an investigating team conducting this type of trial needs to be fully aware of the potential difficulties.

Ethical considerations, as judged perhaps by a local, national or international committee, may also prevent the 'optimal' design being implemented. There are also issues related to patient data confidentiality which may, in the circumstances of a multicentre trial, make synthesis of all the trial data problematical. We address other components of Figure 1.3 in later sections of the book.

1.5 Influencing clinical practice

As we have indicated, an important consideration at the design stage of a trial is to consider whether, if the new treatment proves effective, the trial will be reliable enough in itself to convince clinical teams not associated with the trial of the findings. Importantly, if a benefit is established will this be quickly adopted into national clinical practice? Experience has suggested that all too frequently trials have less impact than they deserve although it is recognised that results that are adopted in practice are likely to be from trials of an appropriate size, conducted by a respected group and have multicentre involvement. Thus, there are considerations, in some sense outside the strict confines of the design, which investigators should heed if their findings are to have the desired impact.

Some basic or administrative things can help reassure the eventual readers of the reliability of the trial results. These include, although some of these may be mandatory, registering the trial itself, involving and informing other clinical colleagues outside the trial team of progress, careful documentation of any serious adverse events, ensuring the trial documentation is complete, establishing procedures for responding to external queries, clarity of the final reporting document in the research literature and seeking avenues for wider dissemination of the trial results.

1.6 History

Probably the single most important contribution to the science of comparative clinical trials was the recognition by Austin Bradford Hill (1897–1991) in the 1940s that patients should be allocated the options under consideration at random so that comparisons should be free from bias. Consequently, the first randomised trial was planned to test the value of a pertussis vaccine to prevent whooping-cough and the results of which were subsequently published by the Medical Research Council

Whooping-Cough Immunization Committee (1951). He later stated: 'The aim of the controlled clinical trial is very simple: it is to ensure that the comparisons we make are as precise, as informative, and as convincing as possible'. This development by itself may not have led to more theoretically based statistical innovation directly, but was the foundation for the science of clinical trials.

Nevertheless, the history of clinical trials research precedes this important development by many years. Thus, clinical trials were mentioned by Avicenna (980–1037) in his *The Canon of Medicine* (1025) in which he laid down rules for the experimental use and testing of drugs and wrote a precise guide for practical experimentation in the process of discovering and proving the effectiveness of medical drugs and substances. His rules and principles for testing the effectiveness of new drugs and medications are summarised in Figure 1.4, and these still form the basis of modern clinical trials.

One of the most famous clinical trials was that conducted by James Lind (1716–1794) in 1747. He compared the effects of various different acidic substances, ranging from vinegar to cider, on groups of sailors afflicted by scurvy and found that the group who were given oranges and lemons had largely recovered from their scurvy after 6 days. Somewhat later, Frederick Akbar Mahomed (1849–1884) founded the Collective Investigation Record for the British Medical Association. This organisation collated data from physicians practising outside the hospital setting and was an important precursor of modern collaborative clinical trials.

The conduct of clinical trials research is multidisciplinary in nature so that a team effort is always needed from the concept stage, through design, conduct, monitoring and reporting. This collaborative effort has led not only to medical developments in many areas but to developments of a more statistical nature. Thus, for those working in cancer and for whom survival was a key endpoint in the clinical trials the two seminal papers published by Peto, Pike, Armitage, *et al.* (1976, 1977) in the *British Journal of Cancer* marked a new era. These papers provided the template for key items essential to the design, conduct, analysis and reporting of randomised trials with emphasis on

(1) The drug must be free from any extraneous accidental quality.

(2) It must be used on a simple, not a composite, disease.

(3) The drug must be tested with two contrary types of diseases, because sometimes a drug cures one disease by its essential qualities and another by its accidental ones.

(4) The quality of the drug must correspond to the strength of the disease. For example, there are some drugs whose heat is less than the coldness of certain diseases, so that they would have no effect on them.

(5) The time of action must be observed, so that essence and accident are not confused.

(6) The effect of the drug must be seen to occur constantly or in many cases, for if this did not happen, it was an accidental effect.

(7) The experimentation must be done with the human body, for testing a drug on a lion or a horse might not prove anything about its effect on man.

Figure 1.4 Avicenna's rules for the experimental use and testing of drugs

those requiring prolonged observation of each patient. In particular, these papers described the Kaplan and Meier (1958) estimate of the survival curve, logrank test and the stratified logrank test in such detail that any careful investigator could follow the necessary steps. A computer program (termed the Oxford program) had also been distributed (some time before the date of the publications themselves), and this allowed the methods suggested by the papers to be implemented. Certainly, for those working in data centres with responsibility for many (often reasonably large) trials, this program facilitated the analysis and helped to ensure that the ideas expressed in these articles were widely disseminated. These papers formed the basic text for those involved in clinical trials for many years and (besides making the ideas accessible to medical statisticians) their role in easing the acceptance of statistical ideas into the clinical community cannot be underestimated.

It should not go unnoticed that David Roxbee Cox was one of the authors of the seminal papers referred to above although his paper describing the proportional-hazards regression model appeared some 4 years earlier (Cox, 1972). His paper was presented at a discussion meeting of the UK Royal Statistical Society and subsequently published in Series B of the Society's journals. This journal deals with the more theoretical aspects of statistical research and does not make easy reading for many statisticians and would not be one to which clinical teams might readily refer. Despite this, this particular paper is probably one of the most cited papers in the medical literature. In brief, the methodology leads to easier analysis of trials with survival time endpoints that include stratification in their design and/or baseline patient characteristics at the time of randomisation which may affect prognosis.

As we have indicated, EBM requires that it is important to critically assess *all* the available evidence about whether an intervention works. Thus, systematic overviews have become a vital component of clinical trial research and are routinely applied *before* launching new trials as a means of confirming the need to carry out a clinical trial or *after* completing trials as a means of synthesising and summarising the current knowledge on the topic of interest. These reviews are the focal interest of the Cochrane Collaboration, and the associated handbook by Higgins, Thomas, Chandler, *et al.* (2019) provides the key to their implementation.

Some developments have not depended on technical advancement (although there are always some) such as the now standard practice of reporting confidence intervals rather than relying solely on p-values at the interpretation stage. Of major importance over this same time period has been the expansion in data processing capabilities and the range of analytical possibilities only made feasible by the amazing development in computer power. Despite many advances, the majority of randomised controlled trials remain simple in design – most often a two-group comparison.

1.7 How do trials arise?

Although the focus of this book is on comparative, or Phase III, trials to establish the relative efficacy of the interventions under test, it should be recognised that these may be preceded by an often extensive research programme starting with the laboratory bench,

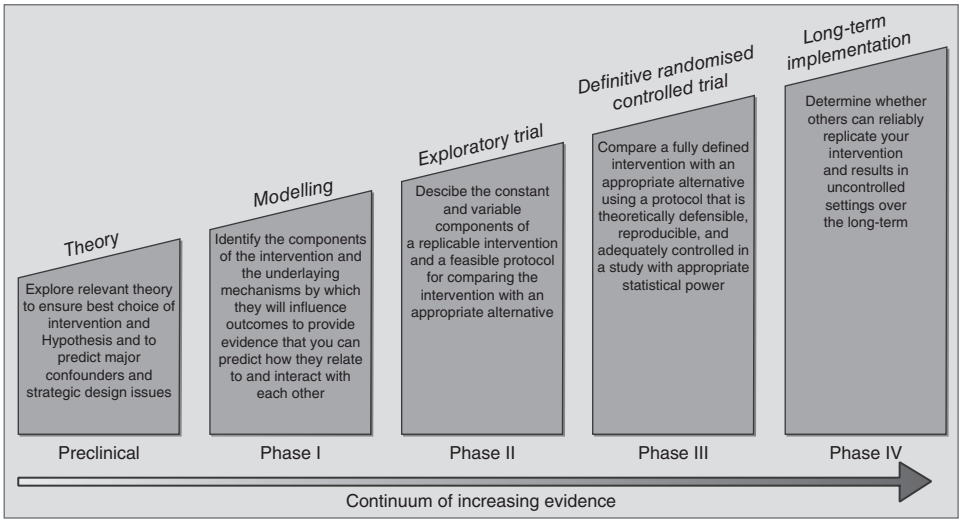

Figure 1.5 Sequential phases of developing randomised controlled trials of complex interventions. Source: Campbell, Fitzpatrick, Haines, *et al.*, (2000).

moving to animal studies and then to early and later stage studies in man. Also once the Phase III stage itself is complete, there may be further studies initiated. Figure 1.5, taken from Campbell, Fitzpatrick, Haines, *et al.* (2000), succinctly summarises the pathway of the whole trial process.

The steps range from studies to determine the pharmacokinetic profile of a drug in healthy volunteers (Preclinical) to establishing the appropriate dosage for use in man (Phase I), then the establishment of indications of activity (Phase II). However, some of these steps may be taken in parallel and even simultaneously in the same subjects.

These early studies are not usually randomised. However, studies conducted by Krishna, Anderson, Bergman, *et al.* (2007) on the effect of the cholesteryl ester transfer protein inhibitor, anacetrapib, on lipoproteins in patients with dyslipidaemia are described by them as 'randomized' and 'phase I'. Randomised they undoubtedly are but their use of the Phase I nomenclature does not have an exact parallel in Figure 1.4. This highlights a difficulty when attempting to categorise trials using such a simple system. One may imagine that there will be clear stages in the development of a bioabsorbable coronary stent. These too will not exactly parallel those of drug development although they may well involve laboratory and animal studies. The single-arm trial of Erbel, Di Mario, Bartunek, *et al.* (2007) may be considered as close to the Phase II type or a feasibility study of Chapter 19.

There are also parallels (although modifications will be necessary) for new approaches to, for example, surgical, radiotherapy or physiotherapy techniques, and combinations of different procedures. They also extend beyond merely therapeutic

trials to planning, for example, trials comparing alternative forms of contraception in women and those evaluating alternative health promotion interventions. However, in some instances, such as in trials comparing educational packages, they may start at the full Phase III stage without involving the earlier phases.

Alternatively, comparative trials may evolve from questions arising in clinical practice and not from a specific development process. Thus, one may wish to compare different surgical timings, at 6 months or at 1 year of age, for reconstructive surgery in infants with cleft palate as is proposed in the trial conducted by Yeow, Young, Chen, *et al.* (2019).

Whatever the pathway, the eventual randomised comparative trial to be conducted is clearly a major event as only when this has been conducted will there be reliable (although not necessarily convincing) evidence of the efficacy of the intervention concerned. In certain situations, often for regulatory purposes, a Phase III trial may be followed by a confirmatory trial asking essentially the same question. In addition, following regulatory approval of a product, so-called Phase IV or postmarketing trials may be initiated with the aim to gain broader experience with using the new product.

1.8 Ethical considerations

For a trial to be ethical, at the time it is designed the ethical review committees will want to be convinced that there is collective uncertainty amongst clinicians as to which treatment is superior or more appropriate for the patients. They will also need to be persuaded that the sample size and other aspects of the study design are such that the trial is likely to provide information sufficient to reduce this uncertainty and thus influence subsequent medical practice if one treatment or the other appears superior.

A clinical trial cannot go forward until the protocol has been through the appropriate ethical review processes, the exact nature of which varies from country to country. These should always include a very thorough review of the scientific aims as well as the more 'subject' oriented concerns to protect those who will be recruited to the trial. In brief terms, this implies that if a trial is not scientifically sound – then it should not be judged as ethically acceptable.

1.9 Regulatory requirements

In addition to the more overtly scientific parts of the clinical trials process on which to focus, there are many regulatory requirements which national and international law obliges a trial team to adhere to. For example, the regulations insist that informed consent is obtained from patients entering trials and on the preservation of personal data confidentiality. These regulations are generally referred to as requirements for Good Clinical Practice (GCP) as is described in ICH E6 (R2) (2016). We will refer to specific aspects of GCP as they arise in the text, but readers are cautioned that the specifics are continually being updated. Principles to guide statisticians working on clinical trials have been laid down by ICH E9 (R1) (2018).

If the trial is seeking regulatory approval of (say) a new drug, then all the associated requirements for approval should be reviewed by the trial team *before*, *during* and *after* the trial protocol is being developed to ensure that all aspects are covered so as to avoid the rejection of the application on what might be a technical detail. For example, there may be a regulatory requirement for some additional animal studies to be conducted before approval can be granted. These requirements are summarised in documents such as those provided by the European Medicines Agency (EMEA) and the US Food and Drug Administration (FDA).

In some circumstances, it is a requirement for regulatory approval that a confirmatory trial is conducted. Such a trial is essentially a repeat of an initial one, perhaps in a different or wider patient group or with a wider group of clinical teams involved, but it must follow the essential features of the predecessor design. Clearly, these details should be cross-checked with the relevant authorities before the protocol is finalised and patients are recruited.

1.10 Focus

As we have illustrated, the size of clinical trials can range from the relatively few to as many as several thousands of subjects being recruited. Consequently, and leaving specific details aside, these will require a range of resources from the relatively modest to the very considerable. It must be emphasised that the size of a clinical trial is determined by the question(s) that are posed, and the resources allocated should reflect the importance of that question. Clearly, a very experienced team is required to launch a large trial but even the design team of an ultimately small sized trial will need access to appropriate personnel including, at a minimum, those with clinical, statistical, data management and organisational skills and often other specialist skills from, for example, pharmacy, pathology and many other specialties. It is important that the design team do not underestimate the scale of the task.

The focus of this book is on the design of (randomised) comparative (usually termed Phase III) trials which are likely to be of relatively modest size. We aim to give clear guidance as to how these may be designed, conducted, (to some extent) analysed and reported. However, it is also important that investigators contributing patients to clinical trials who are perhaps not part of the design team also understand the issues concerned as the very success of the trials depends crucially on their collaboration and understanding of the processes involved.

1.11 Further reading

Although Day (2007) provides a comprehensive list of books about clinical trials the following are particularly useful:

Day S (2007). *Dictionary for Clinical Trials*. (2nd edn). Wiley, Chichester.
Fitzpatrick S (2008a). *Clinical Trial Design*. ICR Publishing, Marlow.

Fitzpatrick S (2008b). *The Clinical Trial Protocol*. ICR Publishing, Marlow.

Girling DJ, Parmar MKB, Stenning SP, Stephens RJ and Stewart LA (2003). *Clinical Trials in Cancer: Principles and Practice*. Oxford University Press, Oxford.

Machin D, Campbell MJ, Tan SB and Tan SH (2018). *Sample Size Tables for Clinical, Laboratory and Epidemiology Studies*. (4th edn). Wiley-Blackwell, Chichester.

Redwood C and Colton T (eds) (2001). *Biostatistics in Clinical Trials*. Wiley, Chichester.

Wang D and Bakhai A (eds) (2006). *Clinical Trials: A Practical Guide to Design, Analysis and Reporting*. Remedica, London

Hints on how to display medical data in tabular and graphical form are given by:

Freeman JV, Walters SJ and Campbell MJ (2008). *How to Display Data*. BMJ Books, Oxford.

For those specifically interested in health-related quality of life issues:

Fayers PM and Machin D (2016). *Quality of Life: The Assessment, Analysis and Interpretation of Patient-reported Outcomes*. (3rd edn). Wiley-Blackwell, Chichester.

For those specifically interested cluster randomised trials at a more technical level:

Campbell MJ and Walters SJ (2014). *How to Design, Analyse and Report Cluster Randomised Trials in Medicine and Health Related Research*. Wiley, Chichester.

For those requiring a wide view of how randomised trials have impacted on clinical practice over a wide range of diseases and conditions:

Machin D, Day S and Green S (eds) (2006). *Textbook of Clinical Trials*. (2nd edn). Wiley, Chichester.