A Concise Guide to Clinical Trials

Allan Hackshaw

WILEY-BLACKWELL
A John Wiley & Sons, Ltd., Publication

BMJ | Books

A Concise Guide to Clinical Trials

Allan Hackshaw

WILEY-BLACKWELL
A John Wiley & Sons, Ltd., Publication

BMJ | Books

This edition first published 2009, © 2009 by Allan Hackshaw

BMJ Books is an imprint of BMJ Publishing Group Limited, used under licence by Blackwell Publishing which was acquired by John Wiley & Sons in February 2007. Blackwell's publishing programme has been merged with Wiley's global Scientific, Technical and Medical business to form Wiley-Blackwell.

Registered office: John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial offices: 9600 Garsington Road, Oxford, OX4 2DQ, UK
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK
111 River Street, Hoboken, NJ 07030-5774, USA

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell

ISBN: 978-1-4051-6774-1

A catalogue record for this book is available from the British Library.

Set in 9.5/12pt Palatino by Aptara® Inc., New Delhi, India
Printed and bound in Singapore

1 2009

Contents

Preface

Clinical trials have revolutionised the way disease is prevented, detected or treated, and early death avoided. They continue to be an expanding area of research. They are central to the work of pharmaceutical companies, which cannot make a claim about a new drug or medical device until there is sufficient evidence on its efficacy. Trials originating from the academic or public sector are more common because they also evaluate existing therapies in different ways, or interventions that do not involve a commercial product.

Many health professionals are expected to conduct their own trials, or to participate in trials by recruiting subjects. They should have a sufficient understanding of the scientific and administrative aspects, including an awareness of the regulations and guidelines associated with clinical trials, which are now more stringent in many countries, making it more difficult to set up and run trials.

This book provides a comprehensive overview of the design, analysis and conduct of trials. It is aimed at health professionals and other researchers, and can be used as an introduction to clinical trials, as a teaching aid, or as a reference guide. No prior knowledge of trial design or conduct is required because the important concepts are presented throughout the chapters. References to each chapter and a reading list are provided for those who wish to learn more. Further details of trial set up and conduct can also be found from country-specific regulatory agencies.

The contents have come about through over 18 years of teaching epidemiology and medical statistics to undergraduates, postgraduates and health professionals, and designing, setting up and analysing clinical studies for a variety of disorders. Sections of this book have been based on successful short courses. This has all helped greatly in determining what researchers need to know, and how to present certain ideas. The book should be an easy-to-read guide to the topic.

I am most grateful to the following people for their helpful comments and advice on the text: Dhiraj Abhyankar, Roisin Cinneide, Hannah Farrant, Christine Godfrey, Nicole Gower, Michael Hughes, Naseem Kabir, Iftekhar Khan, Alicja Rudnicka, and in particular Roger A'Hern. Very special thanks go to Jan Mackie, whose thorough editing was invaluable. And final thanks go to Harald Bauer.

Allan Hackshaw
Deputy Director of the Cancer Research UK & UCL Cancer Trials Centre

Foreword

No one would doubt the importance of clinical trials in the progress and practice of medicine today. They have developed enormously over the last 60 years, and have made significant contributions to our knowledge about the efficacy of new treatments, particularly in quantifying the magnitude of their effects. Crucial in this development was the acceptance, albeit with considerable initial opposition, to randomisation – essentially tossing a coin to determine treatment allocation. Over the past 60 years clinical trials have become highly sophisticated, in their design, conduct, statistical analysis and the processes required before new medicines can be legally sold. They have become expensive and requiring large teams of experts covering pharmacology, mathematics, computing, health economics and epidemiology to mention only a few. The systematic combination of the results from many trials to provide clearer results, in the form of meta-analyses, have themselves developed their own sophistication and importance.

In all this panoply of activity and complexity it is easy to lose sight of the elements that form the basis of good science and practice in the conduct of clinical trials. Allan Hackshaw, in this book, achieves this with great skill. He informs the general reader of the essential elements of clinical trials; how they should be designed, how to calculate the number of people needed for such trials, the different forms of trial design, and importantly the recognition that a randomised clinical trial is not always the right way to obtain an answer to a particular medical question.

As well as dealing with the scientific issues, this book is useful in describing the terminology and procedures used in connection with clinical trials, including explanations of phase I, II, III and IV trials. The book describes the regulations governing the conduct of clinical trials and those that relate to the approval and sale of new medicines – an area that has become extremely complicated, with few people having a grasp of the "whole" picture.

This book educates the general medical and scientific reader on clinical trials without requiring detailed knowledge in any particular area. It provides an up to date overview of clinical trials with commendable clarity.

Professor Sir Nicholas Wald
Director, Wolfson Institute of Environmental & Preventive Medicine
Barts and The London School of Medicine & Dentistry

CHAPTER 1

Fundamental concepts

This chapter provides a brief background to clinical trials, and why they are considered to be the 'gold standard' in health research. This is followed by a summary of the main types of trials, and four key design features. Further details on design and analysis are given in Chapters 3–7.

1.1 What is a clinical trial?

There are two distinct study designs used in health research: observational and experimental (Box 1.1). Observational studies do not intentionally involve intervening in the way individuals live their lives, or how they are treated. However, clinical trials are specifically designed to intervene, and then evaluate some health-related outcome, with one or more of the following objectives:

- to diagnose or detect disease
- to treat an existing disorder
- to prevent disease or early death
- to change behaviour, habits or other lifestyle factors.

Some trials evaluate new drugs or medical devices that will later require a **licence** (or **marketing authorisation**) for human use from a regulatory authority, if a benefit is shown. This allows the treatment to be marketed and routinely available to the public. Other trials are based on therapies that are already licensed, but will be used in different ways, such as a different disease group, or in combination with other treatments.

An **intervention** could be a single **treatment** or **therapy**, namely an administered substance that is injected, swallowed, inhaled or absorbed through the skin; an exposure such as radiotherapy; a surgical technique; or a medical/dental device. A combination of interventions can be referred to as a **regimen**, such as, chemotherapy plus surgery in treating cancer. Other interventions could be educational or behavioural programmes, or dietary changes. Any administered drug or micronutrient that is examined in a clinical trial with the specific purpose of treating, preventing or diagnosing disease is usually referred to as an **Investigational Medicinal Product (IMP)** or **Investigational**

A concise guide to clinical trials, First edition. By A. Hackshaw. Published 2009 by Blackwell Publishing, ISBN: 978-1-4051-6774-1.

Box 1.1 Study designs in health research

Observational

Cross-sectional: compare the proportion of people with the disorder among those who are or are not exposed, at one point in time.

Case-control: take people with and without the disorder now, and compare the proportions that were or were not exposed in the past.

Cohort: take people without the disorder now, and ascertain whether they happen to be exposed or not. Then follow them up, and compare the proportions that develop the disorder in the future, among those who were or were not exposed.

Semi-experimental

Trials with historical controls: give the exposure to people now, and compare the proportion who develop the disorder with the proportion who were not exposed in the past.

Experimental

Randomised controlled trial: randomly allocate people to have the exposure or control now. Then follow them up, and compare the proportions that develop the disorder in the future between the two groups.

An 'exposure' could be a new treatment, and those 'not exposed' or in a control group could have been given standard therapy.

New Drug (IND).[#] An IMP could be a newly developed drug, or one that already is licensed for human use. Most clinical trial regulations that are part of law in several countries cover studies using an IMP, and sometimes medical devices.

Throughout this book, 'intervention', 'treatment' and 'therapy' are used interchangeably. People who take part in a trial are referred to as 'subjects' or 'participants' (if they are healthy individuals), or 'patients' (if they are already ill). They are allocated to trial or intervention arms or groups.

Well-designed clinical trials with a proper statistical analysis provide robust and objective evidence. One of the most important uses of evidence-based medicine is to determine whether a new intervention is more effective than another, or that it has a similar effect, but is safer, cheaper or more convenient to administer. It is therefore essential to have good evidence to decide whether it is appropriate to change practice.

[#] IMP in the European Union, and IND in the United States and Japan.

World Health Organization definition of a clinical trial[1,2]

Any research study that prospectively assigns human participants or groups of humans to one or more health-related interventions to evaluate the effects on health outcomes.

Health outcomes include any biomedical or health-related measures obtained in patients or participants, including pharmacokinetic measures and adverse events.

1.2 Early trials

James Lind, a Scottish naval physician, is regarded as conducting the first clinical trial.[3] During a sea voyage in 1747, he chose 12 sailors with similarly severe cases of scurvy, and examined six treatments, each given to two sailors: cider, diluted sulphuric acid, vinegar, seawater, a mixture of several foods including nutmeg and garlic, and oranges and lemons. They were made to live in the same part of the ship and with the same basic diet. Lind felt it was important to standardise their living conditions to ensure that any change in their disease is unlikely to be due to other factors. After about a week, both sailors given fruit had almost completely recovered, compared to little or no improvement in the other sailors. This dramatic effect led Lind to conclude that eating fruit was essential to curing scurvy, without knowing that it was specifically due to vitamin C. The results of his trial were supported by observations made by other seamen and physicians.

Lind had little doubt about the value of fruit. Two important features of his trial were: a **comparison** between two or more interventions, and an attempt to ensure that the subjects had **similar characteristics**. That the requirement for these two features has not changed is an indication of how important they are to conducting good trials that aim to provide reliable answers.

One key element missing from Lind's trial was the process of **randomisation**, whereby the decision on which intervention a subject receives cannot be influenced by the researcher or subject. An early attempt to do this appeared in a trial on diphtheria in 1898, which used day of admission to allocate patients to the treatments.[4] Those admitted on one day received the standard therapy, and those admitted on the subsequent day received the standard therapy plus a serum treatment. However, some physicians could have admitted patients with mild disease on the day when the serum treatment would be given, and this could bias the results in favour of this treatment. The Medical Research Council trial of streptomycin and tuberculosis in 1948 is regarded as the first to use random numbers.[5] Allocating subjects using a random number list meant that it was not possible to predict what treatment would be given to each patient, thus minimising the possibility of bias in the allocation.

1.3 Why are research studies, such as clinical trials, needed?

Smoking is a cause of lung cancer, and statin therapy is effective in treating coronary heart disease. However, why do some people who have smoked 40 cigarettes a day for life not develop lung cancer, while others who have never smoked a single cigarette do? Why do some patients who have had a heart attack and been given statin therapy have a second attack, while others do not. The answer is that people *vary*. They have different body characteristics (for example, weight, height, blood pressure and blood measurements), different genetic make-up and different lifestyles (for example, diet, exercise, and smoking and alcohol consumption habits). This is all referred to as **variability** or **natural variation.** People react to the same exposure or treatment in different ways; what may affect one person may not affect another. When a new intervention is evaluated, it is essential to consider if the observed responses are consistent with this natural variation, or whether there really is a treatment effect. Variability needs to be allowed for in order to judge how much of the difference seen at the end of a trial is due to natural variation (i.e. chance), and how much is due to the action of the new intervention. The more variability there is, the harder it is to see if a new treatment is effective. Detecting and measuring the effect of a new intervention in the setting of natural variation is the principal concern of medical statistics, used to design and analyse research studies.

Before describing the main design features of clinical trials, it is worth considering other types of studies that can assess the effectiveness of an intervention, and their limitations.

1.4 Alternatives to clinical trials

Evaluating a new intervention requires comparing it with another. This can be done using a randomised clinical trial (RCT), observational study or trial with historical controls (Box 1.1). Although observational studies need to be interpreted carefully with regard to the design features and other influential factors, their results could be consistent with those from an RCT. For example, a review of 20 observational studies indicated that giving a flu vaccine to the elderly could halve the risk of developing respiratory and flu-like symptoms.[6] Practically the same effect was found in a large RCT.[7]

One of the main limitations of observational studies is that the treatment effect could be larger than that found in RCTs or, worse still, a treatment effect is found but RCTs show either no evidence of an effect, or that the intervention is worse. An example of the latter is β-carotene intake and cardiovascular mortality. Combining the results from six observational studies indicated that people with a high β-carotene intake, by eating lots of fruit and vegetables, had a much lower risk of cardiovascular death than those with a low intake (31% reduction in risk).[8] However, combining the results from four randomised trials showed that a high intake might increase the risk by 12%.[8]

Observational (non-randomised) studies

Observational studies may be useful in evaluating treatments with large effects, although there may still be uncertainty over the actual size of the effect. They can be larger than RCTs and therefore provide more evidence on side-effects, particularly uncommon ones. However, when the treatment effect is small or moderate, there are potential design problems associated with observational studies that make it difficult to establish whether a new intervention is truly effective. These are called **confounding** and **bias**.

Several observational studies have examined the effect of a flu vaccine in preventing flu, respiratory disease or death in elderly individuals. Such a study would involve taking a group of people aged over 60 years, then ascertaining whether each subject had had a flu vaccine or not, and which subsequently developed flu or flu-related illnesses. An example is given in Figure 1.1.[9] The chance of developing flu-like illness was lower in the vaccine group than in the unvaccinated group: 21 versus 33%. But did the flu vaccine really work?

The vaccinated group may be people who *chose* to go to their family doctor and request the vaccine, or the doctor or carer recommended it, perhaps on the basis of a perceived increased risk. Unvaccinated people could include those who refused to be vaccinated when offered. It is therefore possible that people who were vaccinated had different lifestyles and characteristics than unvaccinated people, and it is one or more of these factors that partly or wholly explains the lower flu risk, not the effect of the vaccine.

Assume that vitamin C protects against acquiring flu. If people who choose to have the vaccine also happen to eat much more fruit than those who are unvaccinated, then a difference in flu rates would be observed (Table 1.1). The difference of 5 versus 10% could be due to the difference in the proportion of people who ate fruit (80 versus 15%). This is **confounding**. However, if fruit intake had not been measured, it could be incorrectly concluded that the difference in flu rates is due to one group being vaccinated and the other not.

When the association between an intervention (e.g. flu vaccine) and a disorder (e.g. flu) is examined, a spurious relationship could be created through a third factor, called a **confounder** (e.g. eating fruit). A confounder is correlated

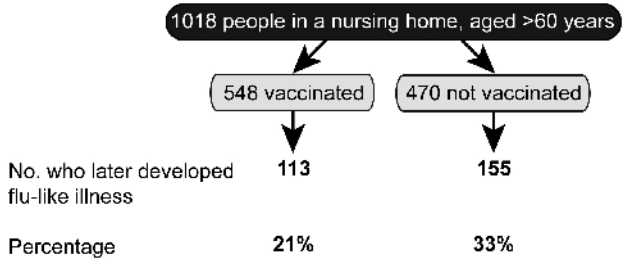

Figure 1.1 Example of an observational study of the flu vaccine.[9]

Table 1.1 Hypothetical observational study of the flu vaccine.

	1000 people aged ≥60 years	
	Vaccinated $N = 200$	*Not vaccinated* $N = 800$
Eat fruit regularly	160 (80%)	120 (15%)
Developed flu 12 months after being vaccinated	10 (5%)	80 (10%)

with both the intervention and the disorder of interest. Confounding factors are often present in observational studies. Even though there are methods of design and analysis that can allow for their effects, there could exist unknown confounders for which no adjustment can be made because they were not measured.

There may also be a **bias**, where the actions of subjects or researchers produce a value of the trial endpoint that is *systematically* under- or over-reported in one trial arm. In the example above, the clinician or carer could deliberately choose fitter people to be vaccinated, believing they would benefit the most. The effect of the vaccine could then be over-estimated, because these particular people may be less likely to acquire the flu than the less fit ones.

Confounding and bias could work together, in that both lead to an under- or over-estimate of the treatment effect, or they could work in opposite directions. It is difficult to separate their effects reliably (Box 1.2). Confounding is sometimes described as a form of bias, since both distort the results. However, it is useful to distinguish them because known confounding factors can be allowed for in a statistical analysis, but it is difficult to do so for bias.

Despite the potential design limitations of observational studies, they can often complement results from randomised trials.[10–14]

Box 1.2 Confounding and bias

- **Confounding** represents the natural relationships between our physical and biochemical characteristics, genetic make-up, and lifestyle and habits that may affect how an individual responds to a treatment. It cannot be removed from a research study, but known confounders can be allowed for in a statistical analysis, and sometimes at the design stage (matched case-control studies).
- **Bias** is usually a design feature of a study that affects how subjects are selected for the study, treated, managed or assessed
- It can be prevented, but human nature often makes this difficult
- It is difficult, sometimes impossible, to allow for bias in a statistical analysis.

Randomisation, within a clinical trial, minimises the effect of confounding and bias on the results

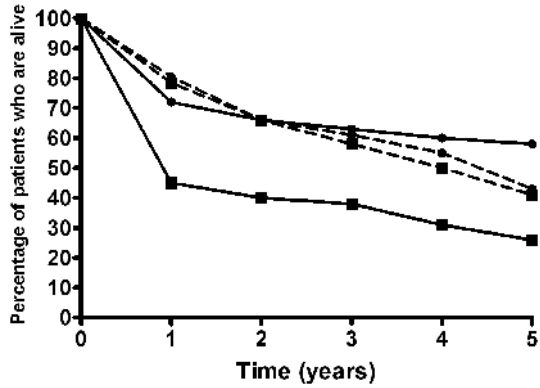

Figure 1.2 Comparison of survival in patients treated with shunt surgery (circles) and medical management (squares). The solid lines are based on a review of five studies, comparing patients treated with surgery at the time of the study, with those treated with medical management in the past. The dashed lines are from a review of eight randomised controlled trials, in which patients were randomly allocated to receive either treatment. The figure is based on information reported in Sacks *et al*.[15]

Historical (non-randomised) controls

Studies using historical controls may be difficult to interpret because they compare a group of patients treated using one therapy now, with those treated using another therapy in the past. The difference in calendar period is likely to have an effect because it may reflect possible differences in patient characteristics, methods of diagnosis or standards of care. Time would be a confounder. In RCTs, subjects in the trial arms are prospectively followed up simultaneously, so changes over time should not matter. The following example illustrates how using historical controls can give the wrong conclusion.

Patients suffering from cirrhosis with oesophageal varices have dilated sub-mucosal veins in the oesophagus. Figure 1.2 shows the summary results on survival in patients treated with surgery (shunt procedures) or medical management.[15] Survival was substantially better in surgical patients in the fives studies that used historical controls, indicated by a large gap between the solid survival curves. However, the eight RCTs showed no evidence of a benefit; the dashed curves are close together. Survival was clearly poorest in the historical control patients, and this could be due to lower standards of care at that time.

1.5 A randomised trial may not always be the best study design

Although a randomised controlled trial is an appropriate design for most interventions, this is not always the case. When planning a study, initial thought should be given to the disorder of interest, the intervention and any information that could affect either how the study is conducted or the results.

The following example illustrates how a randomised trial could be inferior to another design.

The UK National Health Service study on antenatal Down's syndrome screening was conducted between 1996 and 2000.[16] Screening involves measuring several serum markers in the pregnant mother's blood, which are used to identify those with a high risk of carrying an affected foetus. The study aimed to compare the second trimester Quadruple test (four serum markers measured at 15–19 weeks of pregnancy) with the first trimester Combined test (an ultrasound marker and two other serum markers measured at 10–14 weeks). The main outcome measure was the detection rate: the percentage of Down's syndrome pregnancies correctly identified by the screening test. Women classified as high risk by the test would be offered an invasive diagnostic test to confirm or rule out an affected pregnancy.

At first glance, a randomised trial seems like the obvious design. Pregnant women would be randomly allocated to have either the Combined test or the Quadruple test. The detection rates in the two trial arms would then be compared. However, there are two major limitations with this approach:

Sample size. Preliminary studies suggested a detection rate of 85% for the Combined test and 70% for the Quadruple test. To detect this difference requires a sample size of 95 Down's syndrome pregnancies in each arm. The prevalence in the second trimester is about 1.7 per 1000 (0.0017), so 56 000 women would be needed in each arm (95/0.0017), or 112 000 in total. This would be a very large study that may not be feasible in a reasonable timeframe.

Bias. About 25% of Down's syndrome pregnancies miscarry naturally between the first and second trimesters of pregnancy. In a randomised trial there would be an expected 127 cases seen in the first trimester and 95 in the second trimester. The problem is that the Combined test group would include affected foetuses destined to miscarry, while the Quadruple test group has already had these miscarriages excluded, because a woman allocated to have this test but who miscarried at 12 weeks would clearly not be screened in the second trimester. The comparison of the two screening tests would not be comparing like with like, and it can be shown that the detection rate for the Combined Test would be biased upwards.

A better design is an observational study where both screening tests can be compared in the same woman, which is what happened.[16] Women had an ultrasound during the first trimester and gave a blood sample in both trimesters, but the Combined or Quadruple test markers were not measured or examined until the end of the study (no intervention based on these results); women just received the standard second trimester test according to local policy, the result of which was reported and acted upon. This design avoids the miscarriage bias because only Down's syndrome pregnancies during or after the second trimester were known and included in the analysis. The comparison of the Combined and Quadruple tests was thus based on the same group of pregnancies. Furthermore, because each woman had

both tests, a within-person statistical analysis could be performed, and this required only half the number needed compared to a randomised two-arm trial (56 000 instead of 112 000).

1.6 Types of clinical trials

Clinical trials have different objectives. The methods for designing and analysing clinical trials can be applied to experiments on almost any object, for example, animals or cells, as well as humans. They can be broadly categorised into four types (Phase I, II, III or IV), largely depending on the main aim (Box 1.3).

Phase I trials

After a new drug is tested in animal experiments, it is given to humans. Phase I trials are therefore often referred to as 'first in man' studies. They are used to examine the pharmacological actions of the new drug (i.e. how

Box 1.3 Types of trials

Phase I

- First time a new drug or regimen is tested on humans
- Few participants (say <30)
- Primary aims are to find a dose with an acceptable level of safety, and examine the biological and pharmacological effects

Phase II

- Not too large (say 30–70 people)
- Aim is to obtain a *preliminary* estimate of efficacy
- Not designed to determine whether a new treatment works
- Produces data in each of the trial arms, that could be used to design a phase III trial

Phase III

- Must be randomised and with a comparison (control) group
- Relatively large (usually several hundred or thousand people)
- Aim is to provide a *definitive* answer on whether a new treatment is better than the control group, or is similarly effective but there are other advantages

Phase IV

- Relatively large (usually several hundred or thousand people)
- Used to continue to monitor efficacy and safety in the population once the new treatment has been adopted into routine practice.

it is processed in the body), but also to find a dose level that has acceptable side-effects. They may provide early evidence on effectiveness.

Phase I trials are typically small, often less than 30 individuals, and based on healthy volunteers. An exception may be in trials in specialties where the intervention is expected to have side effects, so it is inappropriate to give it to healthy people, but rather those who already have the disorder of interest (e.g. cancer). Subjects are closely monitored. Phase I studies may be conducted in a short space of time, with few recruiting centres, depending on how common the disease is and the type of intervention. There may be several phase I trials, and if the results are favourable, they are used to design a phase II trial. Many new drugs are not investigated further.

Phase II trials

The aim of a phase II study is to obtain a preliminary assessment of efficacy in a group of subjects that is not large, say less than 100 and often around 50. These trials can be conducted relatively quickly, without spending too many resources (participants, time and money) on something that may not work. As in phase I studies, participants are closely monitored for safety.

A phase II study could have several new treatments to examine. There could also be a control arm in which subjects are given standard therapy, because the disease of interest is relatively uncommon, so there is uncertainty over the effect of the standard therapy. If the results are positive, the data in each arm are used to design a randomised phase III trial, for example estimating sample size. When there is more than one intervention, it is best, though not absolutely necessary, to randomise subjects to the trial groups. The advantages of randomising are given on page 12. A randomised phase II study could also provide information on the feasibility of a subsequent phase III trial, such as how willing subjects are to be randomised.

Phase III trials

A phase III trial is commonly referred to as a **randomised controlled trial (RCT)**. Subjects must be **randomly allocated** to the intervention groups, and there must be a **control (comparison)**. The aim is to provide a definitive answer on whether a new intervention is better than the control, or sometimes whether they have a similar effect. Sometimes, there are more than two new interventions. Phase III studies are often large, involving several hundred or thousand people. Results should be precise and robust enough to persuade health professionals to change practice. The larger the trial, the more reliable the conclusions. The size of these trials, and the need for several recruiting centres, mean that they can take several years to complete.

There is sometimes a misunderstanding that a randomised phase II trial is a quick randomised phase III trial, but they have quite different purposes. A randomised phase II study is not usually designed for a direct statistical comparison of the trial endpoint between the two interventions, and this is reflected in the smaller sample size. Therefore, the results cannot be used to make a reliable conclusion on whether the new intervention is better.

However, a phase III trial is designed for a direct comparison, allowing a full evaluation of the new intervention and, usually, a definitive conclusion.[#]

Phase III trials should be designed and conducted to a high standard, with precise quantitative results on efficacy and safety. This can be particularly important for pharmaceutical companies who wish to obtain a marketing licence from a regulatory agency for a new drug or medical device, which normally requires extensive data before a licence is granted. Trials used in this way can be referred to as **pivotal trials**.

Phase IV trials

These are sometimes referred to as **post-marketing** or **surveillance** studies. Once a new treatment has been evaluated using a phase III trial and adopted into clinical practice, some organisations (usually the pharmaceutical industry) continue to monitor the efficacy and safety of the new intervention. Because several thousand people could be included, phase IV studies may be useful in identifying uncommon adverse effects not seen in the preceding phase III trials. They are also based on subjects in the general target population, rather than the selected group of subjects who agree to participate in a phase III trial. However, phase IV studies are not as common as the other trial types, particularly in the academic or public sector. Comparisons can sometimes only be made with historical controls or groups of people (non-users of the new drug) who are likely to have different characteristics. Because of this, phase IV studies are not discussed in further detail in this book, though the methods of analysis for phase III trials can be used.

1.7 Four key design features

The study population of all types of clinical trials must be defined by the **inclusion and exclusion criteria**. The strength of randomised phase II and III trials comes from three further design features: **control**, **randomisation** and **blinding**.

Inclusion and exclusion criteria

It is necessary to specify which participants are recruited. This is done using a set of **inclusion and exclusion** criteria (or **eligibility list**), which each subject has to fulfil before entry. Every trial will have its own criteria depending on the objectives, and this may include an age range, having no serious co-morbid conditions, the ability to obtain consent, and that subjects have not previously taken the trial treatment. They should have unambiguous definitions to make recruiting subjects easier.

[#]Some researchers design a study as if it were a phase III trial, but using a one-sided test with a permissive level of statistical significance $\geq 10\%$ (see Chapter 5) and usually a surrogate endpoint (see Chapter 2). It is however referred to as a randomised phase II trial. The description of randomised phase II studies given in this book is the one preferred here.

Table 1.2 Hypothetical example of inclusion and exclusion criteria for a trial of a new drug for preventing stroke.

Narrow set of criteria	
Inclusion	*Exclusion*
Male	History of heart disease or stroke
Age 50 to 55 years	History of cancer
Never-smoker	Female
	Ex and current smokers
	Unable to give informed consent
	Family history of heart disease
	Average alcohol intake <2 units per day
Wide set of criteria	
Inclusion	*Exclusion*
Male or female	Unable to give informed consent
Age 45 to 85 years	

Determining the eligibility criteria necessitates balancing the advantages and disadvantages of having a highly selected group against those associated with including a wide variety of subjects. Having many criteria which are narrow (Table 1.2), produces a group in which there should be relatively little variability. Subjects are more likely to respond to the treatment in a similar manner, and this makes it easier to detect an effect if it exists, especially if the effect is small or moderate. However, the trial results may only apply to a small proportion of the population, and so may not be easily generalisable. A trial with few criteria, that are wide (Table 1.2), will have a more general application, but the amount of variability is expected to be high. This could make it more difficult to show that the treatment is effective. When there is much variability, sometimes only large effects can be detected easily.

Control group

The outcome of subjects given the new intervention is always compared with that in a group who are not receiving the new intervention. A **control** group normally receives the current standard of care, no intervention or placebo (see Blinding below). Treatment effects from randomised trials are therefore always relative. The choice of the control intervention depends on the availability of alternative treatments. When an established treatment exists, it is unethical to give a placebo instead because this deprives some subjects of a known health benefit.

Randomisation

In order to attribute a difference in outcome between two trial arms to the new treatment being tested, the characteristics of people should be similar between the groups. In the hypothetical example of the flu vaccine (Table 1.1),

Box 1.4 Randomisation

- Randomly allocating subjects produces groups that are as similar as possible with regard to all characteristics except the trial interventions
- The only systematic difference between the two arms should be the treatment given
- Therefore, any differences in results observed at the end of the trial should be due to the effect of the new treatment, and not to any other factors (or differences in characteristics have not spuriously produced a treatment effect, when the aim is to show that the interventions have a similar effect).

the difference in flu risk at the end of the trial could be due to the difference in those who ate fruit regularly (**confounding**), not the vaccine. Randomly allocating patients to the trial arms means that any difference in outcome at the end of the trial should be due to the new treatment being tested, and not any other factor (Box 1.4).

Randomisation is a process for allocating subjects between the different trial interventions. Each subject has the same chance of being allocated to any group, which ensures similarity in characteristics between the arms. This minimises the effect of both known and unknown confounders, and thus has a distinct advantage over observational studies in which statistical adjustments can only be made for known confounders. Although randomisation is designed to produce groups with similar characteristics, there will always be small differences because of chance variation. Randomisation cannot produce *identical* groups.

Randomisation also minimises **bias**. If either the researcher or trial subject is allowed to decide which intervention is allocated, then subjects with a certain characteristic, for example, those who are younger or with less severe disease, could be over-represented in one of the trial arms. This could produce a bias which makes the new intervention look effective when it really is not, or over-estimate the treatment effect. **Selection bias** can occur if a choosing a particular subject for the trial is influenced by knowing the next treatment allocation. **Allocation bias** involves giving the trial treatment that the clinician or subject feels might be most beneficial. Sometimes, the researcher has access to the list of randomisations from which the next allocation can be seen, possibly creating allocation bias. This can be avoided if randomisation is done through a central office (for example, a clinical trials unit) or a computer system, because the researcher has no control over either process (called **allocation concealment**).

Blinding

The randomisation process minimises the potential for bias, but the benefit could be greater if the trial intervention given to each subject is concealed. Subjects or researchers may have expectations associated with a particular treatment, and knowing which was given can create bias. This can affect how