Chapter 7
Intervention trials

Intervention trials are the epidemiological studies that most closely resemble the experiments conducted by scientists in the laboratory. The essential and distinguishing feature of such studies lies in the investigator's direct control over the allocation of subjects to study groups. In contrast, in observational studies, the allocation is determined by the subjects themselves and the researchers are just passive observers of what happens.

Intervention trials provide the strongest evidence with which to test hypotheses. However, they are not the most usual study design in epidemiology, mainly because of ethical constraints. It would be unacceptable to allocate people to either be or not be exposed to a substance or to be subjected to a procedure for which there is some suspicion that it may be harmful. It is, however, possible to conduct a trial to test whether removal of such an exposure will decrease subsequent incidence and mortality. Thus, intervention trials in epidemiology are limited to interventions for which there are grounds to believe that there will be a potential benefit to individuals.

7.1 Types of intervention study

Intervention trials consist of trials to prevent disease (field trials) or trials to treat established disease processes (clinical trials).

The objective of a *clinical trial* is to evaluate one or more new treatments for a disease or condition. For instance, a clinical trial may be designed to assess whether a chemotherapeutic agent can prevent recurrence of cancer, increase survival or improve quality of life (Example 7.1). Since clinical trials involve diseased people, they are often carried out in hospitals or other clinical settings where the subjects are treated and followed up for their condition.

> **Example 7.1.** *A total of 474 adult patients with malignant glioma (astrocytoma) grade 3 or 4 were randomized to receive 45 Gy (in 20 fractions over four weeks) or 60 Gy (in 30 fractions over six weeks) of radiotherapy postoperatively. The main objective of the study was to assess whether the higher dose would improve survival (Bleehen & Stenning, 1991).*

In contrast, *field trials* deal with subjects who are disease-free. A field trial involves evaluation of whether an agent or procedure reduces the risk of developing disease among those free from that condition at enrolment. Because these trials involve healthy rather than diseased people, they tend to be logistically more difficult to carry out than clinical trials. They generally

have to be conducted in the 'field' rather than in hospitals or clinics. Moreover, whereas the adverse consequences of a given disease (e.g., disease recurrence, death) may occur with high probability during a relatively short time, typically the risk of contracting a disease among people who are initially free of it is small. This is particularly true for rare diseases such as cancer. Consequently, field trials usually require a greater number of subjects followed up for longer periods than clinical trials.

> *Example 7.2. A randomized trial was carried out among Whitehall (English) civil servants to measure in middle-aged men the health effects of stopping smoking. A total of 1445 male cigarette smokers aged 40–59 years who were at a high risk of developing cardiorespiratory diseases were randomly allocated to intervention (714 men) or normal care (731 men). Those in the intervention group received individual advice on the relation of smoking to health. Most then expressed their wish to stop smoking and received further support over the next 12 months. The two groups were then followed up for twenty years (Rose & Colwell, 1992).*

Field trials can be carried out among *individuals* (as in Example 7.2) or *groups of people* (as in Examples 7.3 and 7.4). In the first case, the unit of allocation to the intervention is the individual, whereas in the second, it is the group. The group may be a household, a block of houses, a school or a whole community. Field trials in which whole communities are the unit of allocation are called *community trials* (Example 7.3).

> *Example 7.3. The Community Intervention Trial for Smoking Cessation (COMMIT) was a multicentre project designed to evaluate a community-wide smoking cessation programme in the USA. This trial began in 1989 in 11 matched pairs of communities. One community of each pair was randomly assigned to receive the smoking cessation programme with the other acting as a control. The intervention was designed to promote smoking cessation by using a wide range of community resources to affect attitudes and policies towards smoking (COMMIT Research Group, 1991).*

> *Example 7.4. A randomized controlled trial was carried out to assess the effectiveness of health education leaflets in reducing the incidence of sunburn (one of the known risk factors for malignant melanoma of the skin) among British holiday-makers. The study population comprised holiday-makers travelling to warmer countries on flights from Manchester airport in the United Kingdom during August 1993. The unit of study was the flight. Flights were randomly allocated to either receive the leaflets (intervention group) or not to (control group) (Dey et al., 1995).*

There are various reasons for selecting groups rather than individuals as the study unit. Many interventions are impossible to assign at an individual level. Environmental interventions such as water fluoridation or improvement of air quality can be conducted only at a group level. Most health education interventions also fall into this category. For instance, the intervention (i.e., the smoking cessation programme) in Example 7.3 was aimed primarily at the community rather than the individual; thus, it was appropriate to choose the community as the study unit. It may also be logistically easier to conduct the trial among groups of people than among individuals. In Example 7.4, for instance, it was much easier to allocate flights to either the intervention group or the control group than it would have been to allocate individuals. By allocating flights, it was also possible to minimize the potential for 'contamination', that is, the possibility that people in the control group would end up having access to the leaflets. Such contamination would have made the two groups more alike and, consequently, would have decreased the ability of the trial to reveal any true effect that the intervention might have had (see Section 7.10).

7.2   Formulation of the study objectives

The main objectives of an intervention study should be clearly specified before its start. They should include a concise, but detailed, description of the intervention to be evaluated, the outcome(s) of interest and the population in which the study will be conducted. For example, it is not enough just to state that the objective of a trial is 'to assess whether administration of tamoxifen prevents the development of breast cancer in women'. It is necessary to define exactly the target population. For instance, does it include all women or only those at high risk of developing the disease? Which age-groups will be included? The intervention also needs to be specified in terms of dose, frequency of administration and duration. In addition, it is necessary to decide whether the comparison group will be given a placebo or nothing at all. The outcome(s) of interest and the procedures used to measure them should also be clearly stated.

It is important to decide at this stage whether the intent of the study is primarily *scientific* (explanatory) or *pragmatic* (Schwartz & Lellouch, 1967). If primarily scientific, the trial should be carried out under ideal conditions, so that it will be possible to establish the maximum benefit that the intervention can achieve. It is sensible in these circumstances to conduct the trial among special groups of people (such as volunteers) so as to ensure a high level of compliance. Pragmatic trials, by contrast, assess whether the intervention works when introduced into a public health or clinical setting, i.e., in real-life conditions. In these studies, the true effect of the intervention is likely to be diluted, among other things, by low levels of compliance.

7.3   Ethical issues

In observational studies, it is the investigator's responsibility to maintain the confidentiality of the data provided by the study subjects and to ensure

that the procedures used to measure the exposures and the outcomes of interest do not involve unacceptable levels of discomfort, stress or risk for the participants.

In intervention trials, however, the situation is different. Researchers are no longer simply observing what happens to the study subjects. Since the investigator is deliberately intervening, ethical considerations are more important than in any other type of epidemiological study. Intervention trials are ethically justified only in a situation of uncertainty, when there is genuine doubt concerning the value of a new intervention in terms of its benefits and risks. The researcher must have some evidence that it may be of benefit, for instance, from laboratory and animal studies, or from observational epidemiological studies. Otherwise, there would be no justification for conducting a trial.

Unfortunately, many medical interventions have never been properly evaluated in well conducted intervention trials. For instance, radical mastectomy was used for more than a hundred years as the standard form of treatment for early breast cancer. It was not until the late 1970s, when clinical trials were finally conducted, that this form of treatment was replaced by more conservative types of breast surgery. The clinical trials revealed that there were no differences in recurrence or survival between patients who underwent radical mastectomy and other (more conservative) types of surgery (Veronesi *et al.*, 1981; Fisher *et al.*, 1985). Thus, women with early breast cancer were unnecessarily subjected to a very mutilating form of surgery for decades because clinicians were convinced that it would have been unethical to deprive women of the standard form of therapy. The lesson from this, and many other examples, is that it is best to conduct a trial when any agent or procedure is first introduced rather than after it has gained widespread acceptance and becomes considered standard practice. Failure to carry out a proper trial, when it is needed and feasible, may also be unethical.

Whether a study is considered to be ethical or unethical is a subjective judgement based on cultural norms, which vary from society to society and over time. A useful reference with proposed guidelines for research involving human subjects is that published by the Council for International Organizations of Medical Sciences (CIOMS) and the World Health Organization (WHO) (1993).

7.4 Target and experimental populations

The *target population* is the general group to whom the investigators expect the results of the trial to be applicable. A trial may concern all human beings, if it is believed that the intervention to be assessed is of potential benefit to everyone, or only certain subgroups of the population, such as women or smokers. Thus, the target population represents the scope of the public health impact of the intervention.
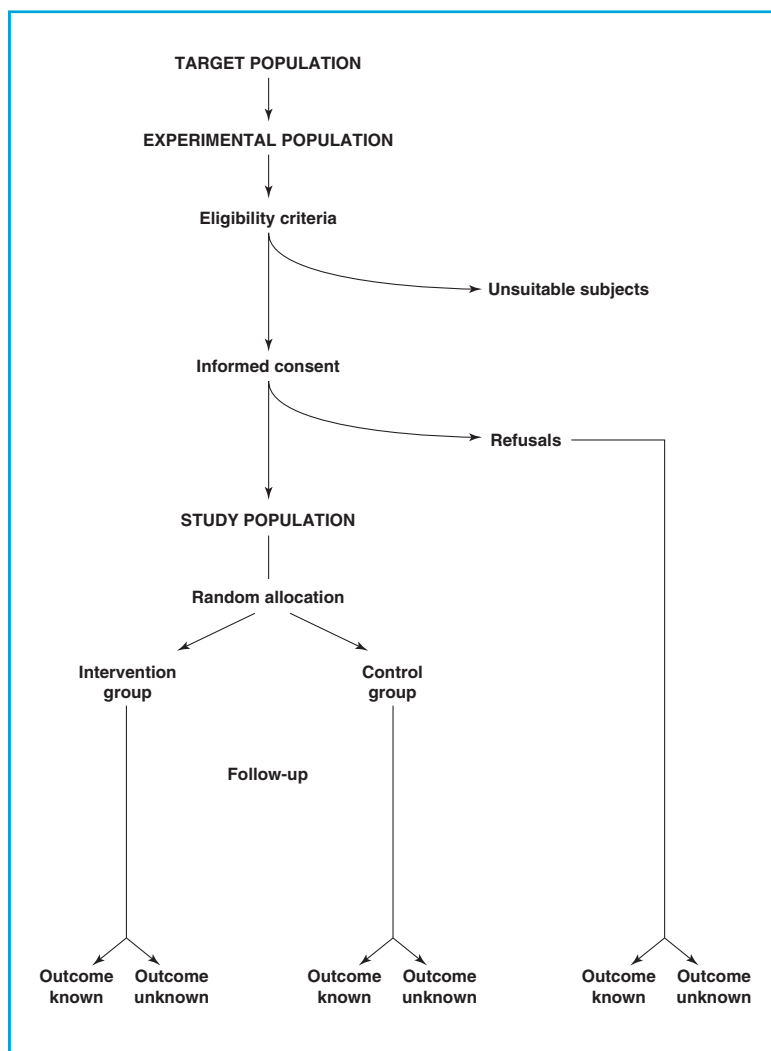
Once the target population has been defined, one needs to select the actual population in which the study will be carried out (Figure 7.1). The choice of this *experimental population* depends on a number of issues. First, it should

not differ from the target population in such a way that generalization to the latter is impossible, although this may be sacrificed in certain circumstances. For example, intervention studies are sometimes carried out among special groups such as volunteers to ensure good compliance or to facilitate the logistics. These trials are useful to evaluate the potential effect of a new intervention, even though it may be difficult to extrapolate the results to the target population. Second, it is essential to determine whether the proposed experimental population is sufficiently large to achieve the sample size necessary for the trial (see Chapter 15). Third, it is important to choose an experimental population that will experience a sufficient number of the outcomes of interest to permit meaningful comparisons between various treatments or procedures within a reasonable period of time. Thus, most trials are carried out in populations where the risk of developing the outcome(s) of interest is high. For instance, to assess the potential benefit of a smoking cessation programme, it would make sense to select as the experimental population one with a high prevalence of tobacco use and high incidence of lung cancer.

The selection of the experimental population also depends on logistic factors. The study should be carried out in an area where it will be possible to obtain support from the local authorities or leaders of the community and where it will be possible to obtain complete and accurate follow-up information for the duration of the trial. For instance, conducting a long-term trial among a highly mobile population such as college students or nomads may result in low follow-up, which would compromise the study.

Figure 7.1.
General outline of an intervention trial.

7.5   Eligibility criteria

Eligibility criteria must be clearly defined before the study begins (Figure 7.1). These should specify exactly who can be included in the study. The criteria will vary from study to study, but, in general, should be such as to eliminate subjects who may be put at greater risk by the intervention or who have an underlying condition that could interfere with the assessment. For instance, patients may be excluded if their physical and/or mental condition

is inadequate to permit interview or collection of biological specimens. It is also usual to exclude pregnant women and women of childbearing age if there is any possibility, however minimal, that the intervention may be harmful to the fetus.

Once the eligibility criteria have been defined, it is possible to establish who are the eligible individuals in the experimental population. Sometimes, it may be necessary to carry out a baseline survey to identify eligible individuals, as in Example 7.5. In these circumstances, subjects must be invited to participate in the baseline survey although they may not fulfil the eligibility criteria for entry into the trial.

Example 7.5. An intervention study was carried out in The Gambia to determine the contribution of bedbugs to hepatitis B transmission. In order to be eligible for the trial, children had to be free from hepatitis B infection at the time of enrolment. All children aged six months to five years living in seven adjacent Mandinka villages were examined by a physician and had a sample of blood taken for serological testing. Only those found to be uninfected were then randomized into the intervention (insecticide spraying of the child's dwelling) or control groups (Mayans et al., 1994).

7.6 Informed consent

The eligible subjects must then be invited to participate in the trial. At this stage, they should be fully informed in simple language of the aims of the study, its procedures, what exactly will be required from them, and of possible risks and benefits. They should also be informed that they will be allocated to either the intervention group or the comparison group and that they may not know which group have they been allocated to until the end of the trial. Subjects should also be assured that their privacy will be respected, that their identity will not be revealed to anyone outside the research team, and that the investigators will not use any information obtained during the study to their detriment (for instance, to compile tax lists). Individuals should be given enough time to consider whether they are willing to participate and they should be allowed to refuse or to withdraw their participation at any time without any negative consequences to them.

If the subjects, provided with this information, still decide to participate in the study, they are said to have given their *informed consent*. In many countries, ethical committees and grant-giving bodies require that the participants sign a consent form (sometimes in the presence of a witness). This may be difficult to accomplish and of relatively little meaning in populations with low levels of literacy.

All efforts should be made to try to explain the nature of the study in a way that the individuals can understand and that it is appropriate to their cultural values and norms. What is appropriate in the 'western' world is not necessarily appropriate in other cultural settings. For example, in some soci-

eties, decisions about participation in a study may be taken at a communal rather than an individual level. Thus, permission to conduct a research project may be obtained through respected community leaders, instead of from individual community members. But even if communal consent is obtained for the study, the investigator still has the responsibility to explain the study procedures and the potential risks and benefits to every single individual who may participate and to ensure that each is aware that he/she is free to refuse to participate or to withdraw from the investigation at any time.

There is a parallel to this situation in western societies. Communal rather than individual consent is generally obtained in community trials, since it is generally impossible to obtain consent from every single member of the communities involved. Consent should be obtained from the local authorities and community leaders. Once these persons have agreed to the communities' participation, it is important for the investigator to inform the community members themselves that they will be participating in a study.

7.7   Study population

Those who are eventually found to be both eligible and willing to enrol in the trial compose the actual *study population* and are often a relatively small and selected subgroup of the experimental population (Figure 7.1). Participants in an intervention study are very likely to differ from non-participants in many ways that may affect the risk of development of the outcomes under investigation. Whether or not the subgroup of participants is representative of the entire experimental population will not affect the validity of the results of a trial conducted among that group. It may, however, affect the ability to generalize those results to either the experimental or the target population.

Example 7.6. The Physicians' Health Study was a randomized, placebo-controlled, double-blind clinical trial conducted in the United States to assess the effects of aspirin on total cardiovascular mortality, and of beta-carotene on cancer incidence. The trial began in 1982, when letters were mailed to 261 248 US male physicians aged 40–84 years asking them to participate. Roughly half of them responded, of whom half again indicated they were willing to participate. Men with a history of cancer (except non-melanoma skin cancer), myocardial infarction, stroke or transient cerebral ischaemia were considered as ineligible. Thus only 33 211 physicians who were both willing and eligible were enrolled in the run-in phase of the trial, lasting from 1 to 6 months, in which they were assigned to active aspirin and beta-carotene placebo treatment. The purpose of this run-in phase was to enhance compliance, since only those physicians who tolerated aspirin and complied with the medication regime were randomized. At the end of this run-in phase, 22 071 were considered eligible for the trial and were randomized (Hennekens et al., 1985, 1996; Stampfer et al., 1985).

In Example 7.6, less than 10% of the original experimental population ultimately entered the trial. Only those physicians who had proven to be good compliers and experienced no adverse effects were randomized, to increase the ability (power) of the study to test the two study hypotheses (see Chapter 15). Although the exclusion of poor compliers limited the generalizability of the results of the trial, it did not affect their validity. In this example, it was far more important to obtain clear answers to the questions being addressed than to try to ensure that it would be possible to extrapolate the results to a wider population.

An effort should be made, however, to obtain baseline data and/or to ascertain outcomes for subjects who are eligible but unwilling to participate. Such information is extremely valuable to assess the presence and extent of differences between participants and non-participants in a particular trial. This will help in judging whether the results among trial participants are generalizable to the target population.

7.8   Choice of the comparison intervention

A key characteristic of an intervention trial is the inclusion of at least one comparison group, against which the effect of the intervention under study is compared. Consideration must be given to what type of intervention the control group should receive. For instance, in a clinical trial, should the control group receive a placebo (a procedure that resembles the new treatment in all respects except that it does not contain the active ingredient(s)), the current best treatment, or nothing at all?

If there is already an established treatment of proven value, it would be unethical to use a placebo. Moreover, in these circumstances, the real pragmatic question is not so much to show whether the new treatment really works but whether it is any better than the existing one. If there is no standard treatment, a placebo is justifiable on the grounds that it makes it possible for the study to be double-blind (see Section 7.11). However, for many interventions it is not possible to devise a suitable placebo. For instance, it is not possible to find a suitable placebo for surgical interventions or for most health education programmes.

7.9   Allocation to the various study groups

Since participants and non-participants may differ in important ways related to the outcome under study, allocation to the various study groups should take place *only after subjects have been determined to be eligible and have expressed willingness to participate*. That is, the non-participants should be eliminated from the pool of potential subjects before allocation to the intervention and control groups is carried out.

7.9.1   Reasons for random allocation

Random allocation is the best method of allocating the study subjects to the different study groups. This method allows chance, and only chance, to determine the assignment of subjects to the various groups. It

is, therefore, the only way of ensuring that any differences in the outcome measures of the trial are due to the effects of the intervention rather than to underlying differences between the groups. Randomization has two major advantages in relation to other methods of allocation:

(1) *Randomization eliminates selection bias on the part of the participants and investigators.*

Randomization eliminates the possibility of any subjective influence in the assignment of individuals to the different study groups. Methods based upon date of birth or date of entry have also been used in some trials, with one intervention being assigned to those who were born (or who report) on even dates and another to those who were born (or who report) on odd dates. The problem with these methods is that it is possible for the investigator to know in advance the group to which a participant will be allocated and this could introduce conscious or unconscious bias into the allocation procedure. An investigator who knows that a particular subject is going to be allocated to a particular intervention may be more or less likely to consider the subject eligible for entry into the study. Randomization can ensure that this does not happen, provided it is done only after subjects have been determined to be eligible and have expressed willingness to participate in the trial.

(2) *Randomization tends to create groups that are comparable in terms of the distribution of known and, more importantly, unknown factors that may influence the outcome.*

Randomization ensures that the distribution of known and, more importantly, of *unknown* confounding variables will be similar in the groups to be compared, provided that the sample size is relatively large. This is unique to experimental studies. Although it is possible in observational studies to take into account the effect of confounders in the analysis, this can only be done for variables which were known or suspected to be confounders at the beginning of the study and for which data were therefore collected (see Chapters 13 and 14). Trials may extend over many years and it is possible that new confounders will become known in the meantime. This would not affect the validity of the results from a randomized trial, however, in which the distribution of any unknown confounding variables would be similar in the study groups *provided that the number of subjects randomized was large*.

In this discussion of randomization, it is worth mentioning that confusion often exists in the use of the expressions '*random allocation*' (or 'random assignment') and '*random sampling*' (or 'random selection'). In this section we are dealing with 'random allocation', namely the process by which subjects are allocated to the study groups in a trial. This constitutes a fundamental principle on which intervention studies are based. Random

selection refers to the process whereby a sample of subjects is selected at random from a larger population. Clinical trials rarely entail random selection; the investigator takes the patients available to him/her, provided they meet the criteria for entry into the study. In field trials, random selection may be used if the experimental population is larger than is required to ensure that the study will have the ability to answer the problem being addressed (that is, that the trial will have adequate power or precision—see Chapter 15). Methods for selecting random samples from a population are discussed in Chapter 10.

7.9.2   Methods of randomization

Various methods can be used to randomize the study subjects to the different study groups. Regardless of the method chosen, it is important to ensure from the earliest stages of the trial that the randomization procedure and the randomization list will be concealed from the persons who are responsible for recruiting the subjects, monitoring the effects of the intervention and assessing the outcomes of the trial.

Simple randomization

Simple randomization is the most elementary method of randomization. It is the equivalent of tossing a coin. However, randomization by tossing a coin should not be used because it cannot be checked or reproduced. The alternative is to use a table of random numbers (or a computer-generated randomization list) (Figure 7.2).

The first step in determining random group assignments is to set up a correspondence between the numbers in the table and the study groups. Let us assume that odd numbers correspond to the control group and even numbers to the new intervention. The second step is to define a convenient way of reading the table of random numbers, for instance, to read down the columns or across the rows.

The third step is to select a starting point, for instance, by closing your eyes and selecting a number with a pin. Once the starting point is established, numbers are then read from the table following the sequence defined in step two. Figure 7.2 is an extract of a table of random numbers (a full table is reproduced in Appendix 7.1). Suppose that the chosen starting point was the one circled in the table and that we have decided that numbers should be read column by column down the page. The first 10 numbers would have been 8, 9, 3, 5, 7, 5, 5, 9, 1, 0. The fourth step is to make the treatment assignments according to the system defined above (Table 7.1).

Random number tables are generated in such a way that each of the digits zero through nine is equally likely to occur. If equal numbers of participants are required in each intervention group, the same number of one-digit numbers should be assigned for each group, even if this means that some digits do not correspond to any group. Thus, in the case of three groups,

Figure 7.2.
Extract from a table of random numbers.

```
67 19 00 71 74    60 47 21 29 68    02 02 37 03 31
02 94 37 34 02    76 70 90 30 86    38 45 94 30 38
79 78 45 04 91    16 92 53 56 16    02 75 50 95 98
87 75 66 81 41    40 01 74 91 62    48 51 84 08 32
34 86 82 53 91    00 52 43 48 85    27 55 26 89 62

11 05 65 09 68    76 83 20 37 90    57 16 00 11 66
52 27 41 14 86    22 98 12 22 08    07 52 74 95 80
07 60 62 93 55    59 33 82 43 90    49 37 38 44 59
04 02 33 31 08    39 54 16 49 36    47 95 93 13 30
01 90 10 75 06    40 78 78 89 62    02 67 74 17 33

92 03 51 59 77    59 56 78 06 83    52 91 05 70 74
61 71 62 99 15    06 51 29 16 93    58 05 77 09 51
73 32 08 11 12    44 95 92 63 16    29 56 24 29 48
42 10 50 67 42    32 17 55 85 74    94 44 67 16 94
26 78 63 06 55    13 08 27 01 50    15 29 39 39 43
```

Patient no.	Sex	Random no.	Allocation
1	F	8	Intervention group
2	M	9	Control group
3	M	3	Control group
4	M	5	Control group
5	M	7	Control group
6	F	5	Control group
7	F	5	Control group
8	F	9	Control group
9	F	1	Control group
10	F	0	Intervention group

Table 7.1.
Example illustrating the use of a table of random numbers to allocate ten subjects to two study groups (see text).

three of the ten one-digit numbers are assigned to each group (e.g., numbers 1, 2, 3 to group A; 4, 5, 6 to group B; and 7, 8, 9 to group C). The remaining number (i.e., zero in our example) in the random tables is ignored and selection moves to the next number.

One of the disadvantages of simple randomization is that it may result in markedly unequal number of subjects being allocated to each group just by chance. For instance, in the above example, only two persons out of ten were assigned to the intervention group. Moreover, simple randomization may also result in the compositions of the different intervention groups being different with respect to factors that may affect the outcome measures in the trial. In the above example, not only was the number of persons allocated to the intervention small but the sex distribution was also quite different in the two groups. This is particularly likely to happen when the total number of subjects in a study is small. For trials involving several hundred participants or more, any such imbalance is likely to be small and can be taken into account in the analysis of the study. In a small trial, imbalance may make the trial more difficult to interpret and, hence, it is advisable to ensure balance by using the randomization methods described below.

Restricted randomization (or blocked randomization)

This method guarantees that the numbers of participants allocated to each study group are equal after every block of so many patients has entered the trial. Suppose that patients are going to be allocated to treatments A and B in such a way that after every fourth subject there are an equal number of participants on each treatment. There are only six possible combinations (permutations) of A and B in blocks of four:

No. Combination
1   AABB
2   ABAB
3   ABBA
4   BBAA
5   BABA
6   BAAB

The combination for a particular block of four patients is chosen at random (by using a table of random numbers as described above) from the six possible (note that in the above example the digits 7, 8, 9 and 0 from the table of random numbers should be ignored). For instance, if the random numbers from the table were 2, 3, 6, 5 (and the blocks were assigned as listed), it would mean that patients 1–4, 5–8, 9–12, 13–16 would receive treatments ABAB, ABBA, BAAB and BABA, respectively. This procedure thus allocates eight patients to group A and eight to group B.

Stratified randomization

When the results of the trial are likely to vary between, say, the sexes or different age-groups, stratified randomization should be used. In this situation, strata or groups are formed and randomization occurs separately for the subjects in each stratum. As subjects become eligible for inclusion in the trial, their appropriate stratum is determined and they receive the next random-number assignment within that stratum. For example, patients may be classified according to sex and age (under 50, and 50 and over), yielding a total of four strata. Within each stratum, each patient will be randomly assigned to either the intervention or the control group. This could be done by using either simple or restricted randomization.

Stratified randomization has the advantage of assuring balance between the groups, at least on the characteristics that determine the strata. The use of this method of randomization in the example described above would ensure that the intervention and the control group would be balanced with respect to sex and age. If stratification had not been employed, the researcher would have run the risk that chance might produce imbalance with regard to these important factors, especially if the number of subjects in the trial was small. The disadvantage with stratified randomization is that it is administratively difficult and cumbersome to execute.

Matched-pair design

A matched-pair design is a special case of stratified randomization in which the strata are each of size 2. Individuals (or communities) are matched into pairs, chosen to be as similar as possible for potential confounding variables such that in the absence of any intervention they would be expected to be at similar risk of the disease under study. The intervention is assigned at random to one member of each pair, with the other member acting as a control.

Matching is unnecessary in large trials, as it is likely that any imbalance between the intervention groups, with respect to risk factors for the occurrence of the outcomes of interest, will tend to even out. Furthermore, it is possible to adjust for any residual imbalance during the data analysis without substantial loss of statistical power.

For small trials, more serious imbalance can arise for which it may be difficult to adjust fully in the analysis. This can be a special problem in trials in which communities are randomized, as it is unusual to be able to

include large numbers of communities (more than 20) in such studies (Example 7.7). Pair-wise matching of similar communities (i.e., communities in which the rates of the disease are likely to be similar in the absence of the interventions to be applied) before the allocation of interventions is likely to be a useful strategy in such situations.

Example 7.7. In the Community Intervention Trial for Smoking Cessation (COMMIT) mentioned in Example 7.3, within each pair, communities were matched on factors such as population demographic characteristics (e.g., population size, age, sex and ethnic composition), degree of urbanization, socioeconomic factors, prevalence of smoking and access to media and health care services. The two paired communities were geographically close enough to permit monitoring by the investigators, but not so close that educational activities in the intervention community would affect the control community. One member of each of the 11 matched pairs was then randomly assigned to receive the health education programme and the other to the control surveillance (COMMIT Research Group, 1991).

7.9.3   Some special experimental designs
Factorial design

   One technique to improve efficiency in intervention trials is to test two or more hypotheses simultaneously in a factorial design. A trial of two hypothesis can utilize a two-by-two factorial design, in which subjects are first randomized to intervention A or B to address one hypothesis, and then within each intervention group there is further randomization to interventions C and D to evaluate a second question (Example 7.8).

Example 7.8. The Physicians' Health Study described in Example 7.6 used a 2×2 factorial design. The physicians were assigned to one of four groups, as shown in Figure 7.3.
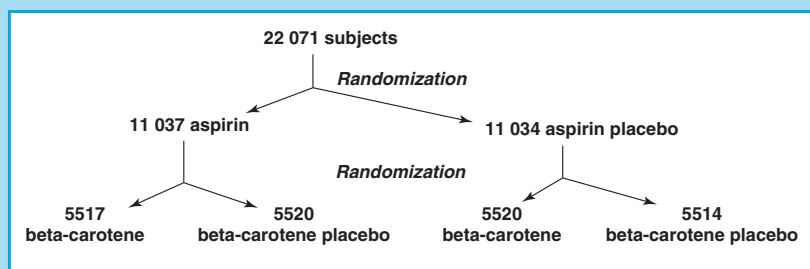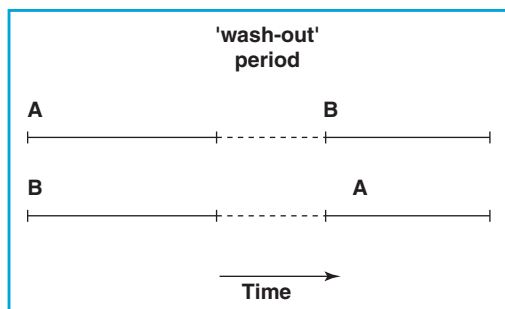

Figure 7.3.
Outline of the factorial design of the Physicians' Health Study.

This design allowed the study of two different questions related to two different diseases: (1) does aspirin reduce the risk of cardiovascular diseases? (2) does beta-carotene reduces the risk of cancer? In addition, it was possible to examine the combined effect of the two drugs on the occurrence of these two diseases (Hennekens et al., 1985; Stampfer et al., 1985).

The principal advantage of the factorial design is its ability to answer two or more questions in a single trial for only a marginal increase in cost. Moreover, the use of a factorial design also allows the assessment of interactions between interventions, which cannot be done in a single-factor study.

Crossover trials

Most trials have a *parallel* design, that is, a group of subjects receives the intervention and another parallel group receives the standard treatment or placebo. In contrast, in *crossover trials* each subject acts as his/her own control by receiving at least two different interventions (e.g., a new drug (treatment A) versus the standard drug (treatment B)) at different times during the trial (Figure 7.4). The order in which each individual receives them (A then B or B then A) should be determined by random allocation

Figure 7.4.
Outline of a crossover trial

(Example 7.9). There should be a 'wash-out' period between each of the interventions to avoid 'carry-over effects' (also called 'spill-over effects'), that is, to ensure that there is no overlap of effects between the first and the second interventions. Consequently, this design is suitable only when neither of the interventions has long-term effects.

The main advantage of crossover trials is that each subject is compared with himself/herself and, therefore, confounding is eliminated from the comparison of the effects of the two treatments (provided that there is no carry-over effect). This design also increases statistical precision in the comparison, because it eliminates inter-subject variability in the outcome response. Hence, fewer subjects are needed than in a corresponding parallel trial.

Crossover trials are used mostly in the early phases of evaluation of new drugs in which their pharmacokinetic properties are investigated in healthy volunteers. They are not appropriate to assess the long-term effects of a treatment, as the nature of the design implies that the treatment period must be limited.

Example 7.9. Thirty-one patients with a diagnosis of metastatic germ-cell tumour and receiving a four-day course of a chemotherapy regimen containing cisplatin were entered in a randomized, double-blind, crossover trial. The objective of this trial was to assess whether oral ondansetron (a serotonin antagonist) plus dexamethasone was more effective than oral ondansetron plus placebo in controlling the emesis associated with chemotherapy. During the first course of chemotherapy, patients were randomly allocated to one of the arms of the trial. Participants were given one of the treatments for eight days. A second course of chemotherapy was given 14 days after the start of the first during which patients crossed over to the alternative anti-emetic regimen (Smith et al., 1991).

7.9.4. Conclusions

It should be emphasized that allocation of subjects to the study groups should be done only after having ascertained that individuals are both eligible and willing to participate. Otherwise subjects who refuse to participate or who withdraw from the study (because the treatment is inappropriate, etc.) will have to be excluded after the randomization, so that the groups may no longer be comparable. This is illustrated in Example 7.10.

Example 7.10. The effect of breast cancer screening on mortality from breast cancer was examined in a randomized trial. Women aged 40–64 years who were members of the Health Insurance Plan of New York were randomly allocated to two groups: an intervention group, whose members were offered four screening examinations (clinical examinations and mammography) at annual intervals; and a control group, who continued to receive their usual medical care. There were about 31 000 women in each group. The groups were very similar with respect to a wide range of demographic and other characteristics. Thirty-five per cent of those offered screening refused (Shapiro, 1977). Table 7.2 shows levels of mortality from causes other than breast cancer for each of the two study groups.

Study group	No. of deaths	Death rate (per 10 000 person-years)
Intervention group	850	57
Screened	421	42
Refused	429	86
Control group	877	58
[a] Data from Shapiro (1977).		

Table 7.2.
Mortality from causes other than breast cancer by the end of the first five years of follow-up.[a]

Table 7.2 shows that there was a large difference in mortality from causes other than breast cancer between women who actually received the intervention (i.e., who were screened) and those in the control group (42 versus 58 per 10 000 pyrs). Since the intervention under study (i.e., breast screening) should not have affected mortality from causes other than breast cancer, the observed difference seem to indicate that the two groups were different in relation to important baseline characteristics. However, if those who refused after randomization were included in the intervention group, as they should be, there is no longer a mortality difference between the two groups (57 versus 58 per 10 000 pyrs).

Random allocation does not necessarily guarantee that the groups will be similar. Discrepancies between the groups may arise just by chance, especially if the number of units being allocated (e.g., individuals, families, communities) is relatively small. Hence, it is essential to collect baseline data on the subjects. These baseline data should include all the variables which are known or thought to affect the outcome(s) of interest and can be used to check the degree of similarity of the groups. If the study

groups differ, statistical techniques can be used that yield results 'adjusted' for any baseline differences (see Chapters 13 and 14).

7.10   Monitoring compliance and side-effects

The problem of achieving and maintaining high compliance is an important issue in the design and conduct of any trial. This is because non-compliance makes the intervention and the comparison groups more alike and, consequently, reduces the ability of the trial to detect any true difference between their outcome measures. A certain degree of non-compliance is acceptable in pragmatic trials, which are aimed at estimating the effectiveness of the intervention in real-life conditions. The aim of scientific trials, however, is to estimate the maximum potential benefit to be derived from the intervention in ideal circumstances, including compliance of 100%. One way of increasing compliance is to use a 'run-in phase' before randomization, as was illustrated in Example 7.6.

Compliance levels must be measured and monitored throughout the study. This can be done by using self-reports. This approach has the disadvantage that it relies exclusively on subjects' judgement and memory. Return of unused medication (e.g., tablets) to the investigators at regular intervals has been used in trials that involve administration of drugs or active substances. However, this method assumes that the subject has taken all the medication that was not returned. Self-reports are sometimes validated against laboratory measurements. Apart from being expensive, these methods also have limited value since they usually measure current and not long-term compliance.

In Example 7.11, the alpha-tocopherol and beta-carotene measurements made three years after entry into the trial were much higher than at baseline in those subjects who were allocated to receive these active ingredients, but changed little in those who were not allocated to receive them (Table 7.3). These findings indicate high levels of compliance.

In most trials, a proportion of participants inevitably become non-compliant for one reason or another (forgetting to take the drugs, developing secondary effects, etc). In such instances, maintaining any level of compliance is preferable to complete non-compliance. Moreover, as will be discussed in Section 7.12, every randomized subject should be included in the primary analysis of any intervention study, so that it is essential to obtain as complete follow-up information as possible on those who have discontinued the intervention programme. Investigators should follow up such individuals for the duration of the trial and obtain information on the relevant outcomes in the same way as for subjects who continue to comply.

Sometimes those who were randomized to one group may choose to obtain the alternative intervention on their own initiative. For instance, those allocated to the control group may adopt the active treatment under study. It is important to minimize this 'contamination' as much as possible. One way is to design the trial in such a way that opportunities of contamination are reduced. For instance, in Example 7.4, flights rather than

Example 7.11. *The Alpha-Tocopherol, Beta-Carotene (ATBC) Cancer Prevention Study was a randomized, double-blind, placebo-controlled trial to determine whether daily supplementation with alpha-tocopherol, beta carotene or both would reduce the incidence of lung and other cancers. A total of 29 133 male smokers aged 50 to 69 years were randomly assigned to one of four regimens: alpha-tocopherol (50 mg per day) alone; beta-carotene (20 mg per day) alone; both alpha-tocopherol and beta-carotene; or placebo. Follow-up lasted for five to eight years. Compliance was assessed by counts of the remaining capsules at each visit, by measurement of serum levels of alpha-tocopherol and beta-carotene after three years of supplementation (Table 7.3), and by measurements in random serum samples throughout the study (Alpha-Tocopherol, Beta Carotene Cancer Prevention Study Group, 1994).*

Treatment received	No. of subjects	Median	20th percentile	80th percentile
Serum alpha-tocopherol levels at baseline				
Alpha-tocopherol	14 472	11.5	9.3	14.2
No alpha-tocopherol	14 469	11.4	9.3	14.1
Serum alpha-tocopherol levels at three years				
Alpha-tocopherol	11 332	17.3	14.3	21.1
No alpha-tocopherol	11 258	12.4	10.2	15.1
Serum beta-carotene levels at baseline				
Beta-carotene	14 460	0.17	0.10	0.29
No beta-carotene	14 460	0.17	0.10	0.29
Serum beta-carotene levels at three years				
Beta-carotene	11 276	3.0	1.6	4.5
No beta-carotene	11 314	0.18	0.10	0.30

[a] Data from Alpha-Tocopherol, Beta Carotene Cancer Prevention Study Group (1994).

Table 7.3.
Serum concentrations (milligrams per litre) of alpha-tocopherol and beta-carotene at baseline and after three years of supplementation by study group.[a]

individuals were chosen as the unit of randomization to minimize contamination. Similarly, in a community trial to evaluate the impact of a smoking cessation programme, it is important that the intervention and the control communities are geographically distinct units with stable populations and no migration between them (as in Example 7.7). Nevertheless, sometimes a certain degree of contamination is inevitable for reasons that are outside the control of the researchers.

In Example 7.12, no difference in the prevalence of oesophageal lesions was found by the end of the trial between the placebo and the treated groups. Laboratory measurements carried out at the time of entry into the study and two months later confirmed that there was a rise in vitamin levels in the active treatment group, but also revealed that the levels of retinol had improved in the placebo group. The change in the placebo group was probably due to better access to fresh fruits and vegetables (Muñoz *et al.*, 1985).

> **Example 7.12.** *A randomized double-blind intervention trial was carried out in Huixian, People's Republic of China, to determine whether combined treatment with retinol, riboflavin and zinc for one year would reduce the prevalence of precancerous lesions of the oesophagus. A total of 610 subjects aged 35–64 years were randomized to the active treatment or placebo. Compliance was very good. At the end of the trial, the prevalence of oesophageal lesions was similar in the two groups: 45.3% in the placebo group and 48.9% in the intervention group (Muñoz et al., 1985).*

A similar contamination problem occurred in the anti-smoking advice trial described in Example 7.2. During the 20-year period that the trial lasted, there was a progressive decline in the prevalence of smoking in the control group (Rose & Colwell, 1992) reflecting a general increase in the awareness of the negative health consequences of smoking. This contamination made the two study groups more alike and reduced the ability of the study to measure the health effects of stopping smoking.

It is also important to monitor any side-effects that might develop. A surveillance mechanism should be set up to allow the breaking of the randomization code if any subject develops serious side-effects. Monitoring side-effects not only is necessary to ensure the safety of the study participants, but also will help to assess the real benefits and hazards of the intervention under study.

7.11 Ascertainment of outcomes

The outcomes of interest should be clearly defined before the start of the trial. The choice of the outcome of interest has important implications for the duration of the trial. Most field trials in cancer epidemiology are aimed at reducing the risk of this disease. Since field trials are conducted among disease-free people, the probability of developing cancer is relatively small and may not be observable for several decades. The problem is less critical in cancer clinical trials, since most of the outcomes of interest (e.g., recurrence or death) tend to occur with a high probability.

One way of shortening the duration of a trial is to select a population that has an increased risk of developing the outcome of interest. For instance, in a field trial to assess the impact of a smoking cessation programme on lung cancer mortality rates, it makes sense to exclude all persons aged under 45 years, since lung cancer is rare at these ages.

Another possibility is to use intermediate endpoints as cancer surrogates, i.e., to use as outcome a biological event that is believed to lie on the causal pathway between exposure and cancer. Studies that use intermediate endpoints are quicker, smaller, and less expensive than studies that use malignancy as the outcome. However, the relevance of the results of the trial with respect to cancer depends on the strength of the association between the intermediate endpoint and the clinical cancer.

> *Example 7.13.* *A randomized, multicentre trial has been set up to test a dietary approach to decreasing the risk of recurrence of polyps of the large bowel. Patients with one or more histologically proven adenomatous polyps, who have had complete removal of polyps at colonoscopy, will be randomly assigned either to usual diet or to nutrition education and counselling aimed at a lifestyle change to a low-fat, high-fibre diet enriched with fruits and vegetables. The trial includes 2000 patients with planned follow-up of four years; colonoscopy will be repeated at years 1 and 4. The major outcome is recurrence of adenomas. The trial is based on the postulate that most large bowel cancers arise from adenomatous polyps. The results of the trial should provide useful evidence about the ability of dietary change to affect recurrence of adenomatous polyps and, hence, to affect indirectly the incidence of large bowel cancer (Chlebowski & Grosvenor, 1994).*

In Example 7.13, only 2000 individuals and a four-year follow-up were required, a substantially smaller number than would have been necessary for a trial having large bowel cancer as the outcome. The underlying assumption in these studies is that the observed relationship between exposure (e.g., diet) and intermediate endpoint (e.g., polyps) reflects a similar one between exposure and cancer *per se*. Clearly, this assumption needs to be validated before any intermediate endpoint can be used as a cancer surrogate (Lippman *et al.*, 1990).

The outcomes should be ascertained in such a way that measurement bias is minimized as far as possible. Blind or masking techniques provide the means for achieving this. When there is no standard intervention to be used in a blind study for comparison with the new intervention, placebos should be employed to maintain blindness. The placebo should be as similar as possible to the intervention itself (with respect to appearance, taste, etc.). Whenever possible, both the patient and the investigators should be unaware of who is assigned to each group until the end of the trial. Such a '*double-blind*' design (both the investigator and the participants are 'blind') eliminates the possibility that knowledge of which intervention an individual is allocated to will affect the way that individual is treated or monitored during the trial, the way the individual responds to the intervention or the way the individual is assessed at the end of the trial. A double-blind trial may not be feasible for the evaluation of programmes involving substantial changes in lifestyle, such as exercise, cigarette smoking or diet, surgical procedures, or drugs with characteristic side-effects. In these circumstances, a '*single-blind*' (the investigator knows to which group a participant belongs but the participant does not or vice-versa) or an unblinded design may be the only possibility.

The more subjective the outcome under study, the greater is the rationale for a double-blind trial. For example, if one deals with extremely subjective responses such as the relief of pain, or the improvement of psychological status, the use of double-blinding is crucial to the validity of the outcome

measurements. When the outcome of a trial is more objective (for example, life or death, or perhaps the level of some substances in the blood or urine), the need for a double-blind trial is, obviously, less important.

The main strength of a double-blind design is to eliminate the potential for measurement bias. Of course, a concomitant limitation is that such trials are usually more complex and difficult to conduct. Procedures must be established for immediate 'unblinding' of a participant's physician in the event of serious side-effects or other clinical emergencies in which this information seems essential.

7.12   Analysis

7.12.1 Types of analysis

There are two main approaches to the analysis of a trial according to who should or should not be included. The '*intention to treat*' analysis is based on outcomes that occur during the whole follow-up period, in the subjects originally allocated to each group, whether they persisted with their allocated intervention or not. The alternative is the '*on randomized treatment*' analysis, which is confined to the outcomes observed while the subjects were on their allocated treatment. Exclusion of randomized subjects of a trial from the analysis may lead to serious bias that can arise from different levels of participation in the intervention and control groups and from the fact that individuals who withdraw or who were lost to follow-up are usually different from those who participate until the end. 'Intention to treat' analysis is the correct way of analysing the data, involving comparison of the outcomes in all the subjects originally allocated to each group (including those who did not have or who stopped having the specified intervention). This stringent approach may sometimes, however, dilute the true effect of the intervention.

7.12.2 On-going analysis

Analysis of results from a trial as data accumulate is an important way of monitoring its progress. Administrative analyses of the numbers of participants recruited each day or week and of the data collected by different field workers are important for quality control.

An independent data-monitoring committee is often set up in large trials to hold the randomization code of the study and to monitor the results of the trial as they come in, or at fixed intervals during the trial. This committee should have the power to stop further recruitment if there is evidence of a substantial risk of adverse reactions associated with any of the interventions under study. Similarly, if evidence accumulates that one intervention is substantially better than the others (or one is substantially worse), the committee can recommend that the intervention phase of the trial be stopped and all participants be given the better (or less harmful) treatment or intervention. It would be very difficult for the investigators to remain objective and impartial if they had to take these decisions themselves.

Example 7.14. The Beta Carotene and Retinol Efficacy Trial (CARET) is a multicentre, randomized, double-blind, placebo-controlled trial set up in 1983 to assess whether a combination of beta-carotene and retinol (vitamin A) could reduce the incidence of lung cancer in populations at high risk. A total of 14 254 heavy smokers and 4060 workers exposed to asbestos were randomized to the active intervention (beta-carotene and vitamin A) or placebo. The design of the trial stipulated that the administration of the intervention should last until late 1997. In January 1996, however, the data-monitoring committee decided to terminate the intervention because it became apparent that there was a 28% increase (95% confidence interval, 0.4% to 57%; P = 0.02) in the risk of lung cancer in the intervention group compared to the placebo group. Follow-up for additional cases is expected to continue for another five years (Omenn et al., 1996).

In Example 7.14, the intervention phase of the trial was terminated early. Post-intervention follow-up of the study subjects will continue for five years to identify additional lung cancer cases and to assess the long-term effects of the intervention.

7.12.3 Final analysis

The first step in the analysis of a trial is to examine the characteristics of the two (or more) groups at baseline to assess their comparability, determining whether randomization resulted in the formation of comparable and evenly balanced groups (Example 7.15). This comparison should constitute the first table of the results section of a paper.

Statistical tests are frequently carried out to assess whether baseline differences between the study groups are important. A statistical test yields the probability of finding by chance a difference at least as large as the one observed. We know, however, that all the observed differences, regardless of their magnitude, have arisen just by chance since the subjects were randomized. Thus, *statistical tests are superfluous and inappropriate to assess whether the study groups have similar baseline characteristics*.

Examination of the baseline characteristics of the groups can also help to reveal any unknown problems that may have occurred during the randomization procedure. For instance, if the baseline characteristics of the groups turn out to be very dissimilar, the entire randomization procedure should be checked for possible deception by some of those in charge of recruiting the subjects into the trial.

After ascertaining the comparability of the study groups, the investigator must determine whether the intervention was of any value. The two groups are compared and the size of the differences assessed. In general, the main results from a trial can be presented in a table similar to one of those shown in Table 7.5.

Table 7.4.
Distribution of baseline characteristics among heavy smokers and asbestos workers who participated in the CARET trial according to type of intervention received.[a]

Characteristic	Workers exposed to asbestos		Heavy smokers	
	Intervention	Placebo	Intervention	Placebo
No. randomized	2044	2016	7376	6878
Age (yrs)[b]	57±7	57±7	58±5	58±5
Female[c]	0	0	3208 (43)	3081 (45)
Race or ethnic group[c]				
White	1805 (88)	1775 (88)	7000 (95)	6487 (94)
Black	152 (7)	153 (8)	103 (1)	122 (2)
Hispanic	36 (2)	43 (2)	101 (1)	95 (1)
Other/unknown	51 (2)	45 (2)	172 (2)	174 (3)
Smoking status[c]				
Never smoked	68 (3)	64 (3)	0	0
Former smoker	1195 (58)	1175 (58)	2473 (34)	2331 (34)
Current smoker	781 (38)	777 (39)	4903 (66)	4547 (66)
Cigarettes smoked/day[b]				
Former smoker	25±12	25±12	28±11	28±11
Current smoker	24±10	25±10	24±9	24±8
Pack-years of smoking (only former and current smokers)[b]	43±24	42±24	50±21	49±20
Years since quitting smoking (only former smokers)[b]	10±8	10±8	3±2	3±2

[a] Data from Omenn *et al.* (1996)
[b] Mean ± standard deviation.
[c] Number (%).

(a)

	Exposure	
	Intervention	Control
Outcome Yes	a	b
No	c	d

Risk in intervention group (p_1) = $a/(a+c)$
Risk in control group (p_0) = $b/(b+d)$
Risk ratio = p_1/p_0
Risk difference[a] = $p_0 - p_1$

(b)

	Exposure	
	Intervention	Control
Cases	a	b
Person-time at risk	y_1	y_0

Rate in intervention group (r_1) = a/y_1
Rate in control group (r_0) = b/y_0
Rate ratio = r_1/r_0
Rate difference[a] = $r_0 - r_1$

[a] In most trials the intervention is protective and thus it is appropriate to calculate the risk and rate difference as indicated (see Section 5.2.2). If the intervention is associated with an increase in incidence, the risk and rate difference should be calculated as $p_1 - p_0$ or $r_1 - r_0$, respectively.

Table 7.5.
Analysis of an intervention study: (*a*) by risk; (*b*) by rates.

If all or virtually all participants enter and leave the trial at the same time, the risk can be calculated (Table 7.5(*a*)). For example, if the follow-up period is uniformly three years, the three-year risk can be computed for each study group. The two study groups can be compared by calculating risk ratios and risk differences as measures of relative and absolute effect, respectively.

In Example 7.16, practically all participants entered the trial at the same point in time and were followed up until its end. If one assumes that, on average, the deaths in each treatment group occurred at similar points in time, the calculation of risk as a measure of disease occurrence is appropriate. The results from this trial were consistent with the null hypothesis of no treatment-associated difference in the risk of lung cancer, that is, a risk ratio equal to 1 (or a risk difference equal to zero).

Example 7.16. In the Physicians' Health Study described in Examples 7.6 and 7.8, 22 071 US male physicians aged 40 to 84 years were randomized in 1982 to receive one of four treatments: (1) aspirin plus beta-carotene placebo; (2) beta-carotene plus aspirin placebo; (3) both active agents; (4) both placebos. The randomized aspirin complement of the trial was terminated early, in 1988, by the external data-monitoring board because it became apparent that there was a 44% reduction (P < 0.001) in the risk of a first myocardial infarction in those taking aspirin. The randomized beta-carotene component continued uninterrupted until its scheduled termination in 31 December 1995. A total of 11 036 physicians received beta-carotene and 11 035 received beta-carotene placebo and fewer than 1% were lost to follow-up. One of the main aims of this component of the study was to assess whether beta-carotene reduces the incidence of lung cancer (Hennekens et al., 1996). Table 7.6 shows the results.

		Study group	
		Beta-carotene	Placebo
Lung cancer	**Yes**	82	88
	No	10 954	10 947
	Total	11 036	11 035

[a] Data from Hennekens *et al.*, 1996

p_1 = 82/11 036 = 0.00743 = 7.43 per 1000

p_0 = 88/11 035 = 0.00797 = 7.97 per 1000

χ^2 = 0.21, 1 d.f.; P > 0.50

Risk ratio (p_1/p_0) = 0.93

95% confidence interval for the risk ratio = 0.69 to 1.26

Risk difference ($p_0 - p_1$) = 7.97 per 1000 − 7.43 per 1000 = 0.54 per 1000

95% confidence interval for the risk difference = − 1.77 per 1000 to 2.85 per 1000

(Test statistics and confidence intervals were calculated using the formulae given in Appendix 6.1).
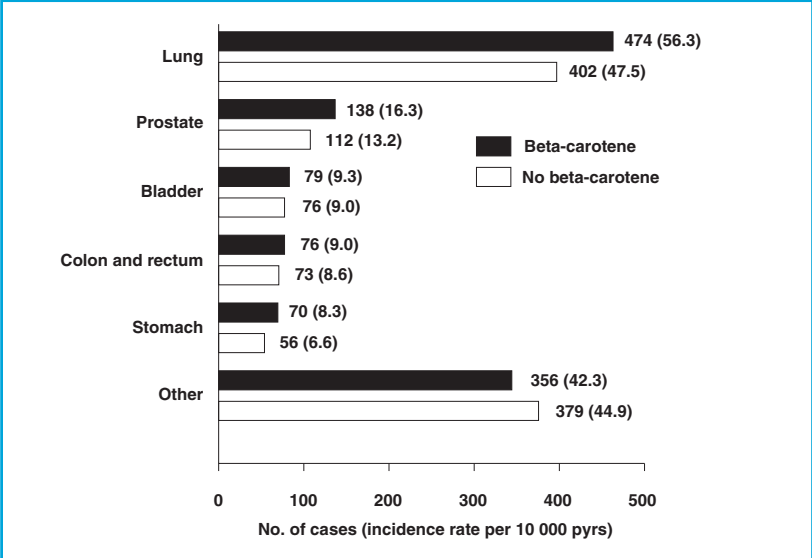
Table 7.6.
Distribution of lung cancer incident cases in the Physicians' Health Study, according to treatment group.[a]

Many intervention trials, however, involve varying periods of follow-up. Recruitment into the trial may take several years and if the follow-up is terminated at a specific point in calendar time, participants will have been observed for different lengths of time. Also, subjects are lost to follow-up or die at different points in time during the study, and consequently they will have been part of the trial for different periods.

Calculation of person-time of observation as the denominator for computation of rates is the method generally used in intervention trials when varying periods of observation (which result from persons entering and leaving the study at different ages and times) have to be taken into account. Results of the trial can be presented as in Table 7.5(*b*), and rate ratios and rate differences calculated as measures of relative and absolute effect, respectively.

Example 7.17. *The objective of the ATBC trial (described in Example 7.11) was to assess whether daily supplementation with alpha-tocopherol, beta-carotene or both would reduce the incidence of lung cancer and other cancers. A total of 29 133 male smokers aged 50 to 69 years from south-western Finland were recruited between 1985 and 1988. Follow-up continued for 5–8 years (median = 6.1), until death or 30 April 1993, when the trial ended (Alpha-Tocopherol, Beta Carotene Cancer Prevention Study Group, 1994). The results by type of treatment received are shown in Figure 7.5.*

Figure 7.5.
Number of cancer cases and incidence rates by site and type of treatment received (reproduced with permission from Alpha-Tocopherol, Beta Carotene Cancer Prevention Study Group, 1994).

The results for lung cancer incidence in relation to beta-carotene supplementation can be presented as shown in Table 7.7.

Table 7.7.
Distribution of lung cancer cases and person-years at risk according to treatment received.[a]

	Study group	
	Beta-carotene	No beta-carotene
Lung cancer cases	474	402
Person-years at risk	84 192	84 632

[a] Data from Alpha-Tocopherol, Beta Carotene Cancer Prevention Study Group, 1994

r_1 = 474/84 192 = 56.3 per 10 000 pyrs

r_0 = 402/84 632 = 47.5 per 10 000 pyrs

χ^2 = 6.30; 1 d.f.; $P \approx$ 0.01

Rate ratio (r_1/r_0) = 1.19

95% confidence intervals for the rate ratio = 1.04 to 1.36

Rate difference (r_1-r_0) = 8.8 per 10 000 pyrs

95% confidence interval for the rate difference = 1.9 to 15.7 per 10 000 pyrs

(Test statistics and confidence intervals were calculated using the formulae given in Appendix 6.1).

In Example 7.17, the duration of follow-up varied from subject to subject. Thus it is more appropriate to calculate person-time at risk and rates as the measure of occurrence of disease. The results of this trial did not support the study hypothesis that beta-carotene reduces the incidence of lung cancer. In fact, they provide evidence that administration of beta-carotene may increase the risk of lung cancer.

Statistical tests and, more importantly, confidence intervals for measures of relative and absolute effect should always be calculated and reported. In Examples 7.16 and 7.17, we used the formulae presented in Appendix 6.1.

The *prevented fraction* (see Section 5.2.2) is another important measure in intervention trials. It measures the proportion of cases of disease that were prevented by the intervention under study among those who received it.

Prevented fraction (%) = 100 × [rate (or risk) difference/rate (or risk) in the unexposed]

If the aim of the trial is to assess the value of a vaccine, this measure is called *vaccine efficacy*. For instance, if the risk of developing a particular disease among those who were vaccinated was 40 per 100 000 and 70 per 100 000 among those not vaccinated, the vaccine efficacy would be

Vaccine efficacy (%) = 100 × [(70 per 100 000 – 40 per 100 000)/70 per 100 000]
= 43%

Thus, 43% of cases could have been prevented among the unvaccinated if they had been vaccinated.

If baseline differences between the study groups need to be taken into account in the analysis, one of the statistical techniques discussed in Chapter 14 should be used.

If we are particularly interested in the distribution of *time until occurrence of the event of interest* (e.g., time from treatment to death or time from treatment to recurrence), as is the case in many clinical trials, the most appropriate approach is *survival analysis*. The techniques used in survival analysis derive from the life-table methods which are discussed in Chapter 12.

Subgroup analysis

It is usual for investigators to perform subgroup analyses to assess whether the intervention has an effect on subgroups of individuals with certain characteristics (e.g., males, elderly people, patients with particular clinical features, etc.). These subgroup analyses raise important problems, however. If the *subgroups are defined according to the baseline characteristics* of the patients, the main concern involves loss of ability of the trial to detect an effect (that is, loss of statistical power or precision (see Chapter

15)), since the results will be based on only a small proportion of the total numbers of randomized subjects. On the other hand, if multiple analyses are performed, some will inevitably achieve 'statistical significance' just by chance. Their interpretation will depend very much on the existence of *a priori* hypotheses based on biological plausibility, existence of supporting evidence from laboratory experiments and from other epidemiological observations.

Analyses performed on *subgroups defined on the basis of individual characteristics which develop after randomization* are of much greater concern, because potential confounding variables will no longer be distributed at random among the subgroups. For instance, analyses restricted to persons who reached a certain serum concentration of the active treatment or who developed a well known secondary effect (e.g., skin yellowing after ingesting beta-carotene) should be treated with extreme caution. Their findings should never be reported as main results of the trial but just as interesting observations that might be worth investigating in specifically designed trials.

7.13 Interpretation

The interpretation of results from a well conducted intervention study should be relatively straightforward, since the two major problems of concern in observational studies, bias and confounding, are greatly reduced by using an experimental design.

This is not to say that trials are exempt from problems. The lung cancer beta-carotene story provides a good illustration of this. Data from three large-scale chemoprevention trials conducted in western countries to assess this question have now been published. The ATBC Cancer Prevention Study (Example 7.11) was set up in Finland to test the hypothesis that a high intake of beta-carotene and alpha-tocopherol reduces the risk of lung cancer. This was a reasonable hypothesis given the substantial evidence available from observational epidemiological studies suggesting that beta-carotene was associated with a lower risk of lung cancer. The results of this trial failed to show any benefit of beta-carotene (or alpha-tocopherol) in the prevention of this malignancy; instead, men who took beta-carotene had an unexpected 'statistically significant' increase in the risk of lung cancer (as we saw in Example 7.17). The authors did consider alternative explanations for this unexpected finding. Confounding could be discarded given the large sample size and the random allocation of subjects to the various study groups. Moreover, the treatment groups were well balanced in relation to relevant baseline characteristics. Since this result was not supported by biological or previous epidemiological evidence, the authors were reluctant to reject the null hypothesis of no effect. They stated at the end of their paper:

"In summary, we found no overall reduction in the incidence of lung cancer or in mortality due to this disease among male smokers who

received dietary supplementation with alpha-tocopherol, beta-carotene, or both in this large trial in Finland. The results of this study raise the possibility that these substances may have harmful effects as well as beneficial effects. Longer observation of the participants in this trial and data from other studies of people at normal or high risk for cancer will be required to determine the full spectrum of effects of these agents. Public health recommendations about supplementation with these micronutrients would be premature at this time" (Alpha-Tocopherol, Beta Carotene Cancer Prevention Study Group, 1994).

Results from two other trials were subsequently published. The active-intervention phase of the CARET trial (Example 7.14) was terminated early because its results confirmed the unexpected increase of lung cancer risk among those who took beta-carotene reported by the ATBC trial. There was again no obvious explanation for this unexpected finding. The Physicians' Health Study (Example 7.8) had a much longer follow-up (12 years) than the other two trials (average of 6 and 4 years for the ATBC and CARET trials, respectively). Its results were consistent with the null hypothesis of no effect of beta-carotene on the risk of lung cancer; in other words, they did not provide evidence of either a beneficial or a harmful effect of beta-carotene. Thus, the lung cancer beta-carotene story shows that results from a single trial should not be considered in isolation.

The results of a trial cannot be translated directly into public health decisions. Other factors that need to be taken into account include issues such as generalizability of the results to different populations, acceptability of the intervention, feasibility, costs, available resources and competing public health priorities. Furthermore, the overall impact of an intervention in a particular population depends not only on the magnitude of the effect of the intervention on the risk of developing a particular condition, but also on the frequency (and severity) of the condition in the population. This issue is further discussed in Chapter 16.

Further reading

* The book by Smith & Morrow (1996) provides a comprehensive coverage of the design, implementation and monitoring of field trials, with particular emphasis on practical aspects. Although the focus is on developing countries, most of the issues discussed in this book are also relevant to developed countries.

* The book by Pocock (1983) provides a good and accessible reference for those interested in the design, analysis and interpretation of clinical trials.

* A short review of methodological issues in design, analysis and interpretation of cancer clinical trials can be found in two papers by Peto *et al.* (1976, 1977).

Box 7.1. Key issues

• Intervention trials are characterized by the fact that investigators are responsible for allocating subjects to the different study groups.

• The main advantages of this type of study are:

　1. *Random allocation* of subjects ensures that allocation of subjects to the different study groups is unaffected by selection bias.

　2. *Random allocation* ensures that the groups are well balanced in relation to known and, more importantly, unknown factors that may affect the outcome(s) of the study (provided the study is sufficiently large).

　3. If the allocation is *double-blind*, measurement bias is also minimized.

　4. *Multiple outcomes* can be studied for any one intervention.

　5. Incidence of disease can be measured in the various study groups.

• The main disadvantages of this type of study are:

　1. Intervention trials, particularly field trials, are large enterprises. They are very expensive and time-consuming.

　2. They may raise important ethical problems.

　3. It may be difficult to ensure compliance and avoid contamination throughout the trial, particularly in trials of long duration.

Appendix 7.1
Table of random numbers

16 22 77 94 39	49 54 43 54 82	17 37 93 23 78	87 35 20 96 43	84 26 34 91 64	
84 42 17 53 31	57 24 55 06 88	77 04 74 47 67	21 76 33 50 25	83 92 12 06 76	
63 01 63 78 59	16 95 55 67 19	98 10 50 71 75	12 86 73 58 07	44 39 52 38 79	
33 21 12 34 29	78 64 56 07 82	52 42 07 44 38	15 51 00 13 42	99 66 02 79 54	
57 60 86 32 44	09 47 27 96 54	49 17 46 09 62	90 52 84 77 27	08 02 73 43 28	
18 18 07 92 46	44 17 16 58 09	79 83 86 19 62	06 76 50 03 10	55 23 64 05 05	
26 62 38 97 75	84 16 07 44 99	83 11 46 32 24	20 14 85 88 45	10 93 72 88 71	
23 42 40 64 74	82 97 77 77 81	07 45 32 14 08	32 98 94 07 72	93 85 79 10 75	
52 36 28 19 95	50 92 26 11 97	00 56 76 31 38	80 22 02 53 53	86 60 42 04 53	
37 85 94 35 12	83 39 50 08 30	42 34 07 96 88	54 42 06 87 98	35 85 29 48 39	
70 29 17 12 13	40 33 20 38 26	13 89 51 03 74	17 76 37 13 04	07 74 21 19 30	
56 62 18 37 35	96 83 50 87 75	97 12 25 93 47	70 33 24 03 54	97 77 46 44 80	
99 49 57 22 77	88 42 95 45 72	16 64 36 16 00	04 43 18 66 79	94 77 24 21 90	
16 08 15 04 72	33 27 14 34 09	45 59 34 68 49	12 72 07 34 45	99 27 72 95 14	
31 16 93 32 43	50 27 89 87 19	20 15 37 00 49	52 85 66 60 44	38 68 88 11 80	
68 34 30 13 70	55 74 30 77 40	44 22 78 84 26	04 33 46 09 52	68 07 97 06 57	
74 57 25 65 76	59 29 97 68 60	71 91 38 67 54	13 58 18 24 76	15 54 55 95 52	
27 42 37 86 53	48 55 90 65 72	96 57 69 36 10	96 46 92 42 45	97 60 49 04 91	
00 39 68 29 61	66 37 32 20 30	77 84 57 03 29	10 45 65 04 26	11 04 96 67 24	
29 94 98 94 24	68 49 69 10 82	53 75 91 93 30	34 25 20 57 27	40 48 73 51 92	
16 90 82 66 59	83 62 64 11 12	67 19 00 71 74	60 47 21 29 68	02 02 37 03 31	
11 27 94 75 06	06 09 19 74 66	02 94 37 34 02	76 70 90 30 86	38 45 94 30 38	
35 24 10 16 20	33 32 51 26 38	79 78 45 04 91	16 92 53 56 16	02 75 50 95 98	
38 23 16 86 38	42 38 97 01 50	87 75 66 81 41	40 01 74 91 62	48 51 84 08 32	
31 96 25 91 47	96 44 33 49 13	34 86 82 53 91	00 52 43 48 85	27 55 26 89 62	
56 67 40 67 14	64 05 71 95 86	11 05 65 09 68	76 83 20 37 90	57 16 00 11 66	
14 90 84 45 11	75 73 88 05 90	52 27 41 14 86	22 98 12 22 08	07 52 74 95 80	
68 05 51 18 00	33 96 02 75 19	07 60 62 93 55	59 33 82 43 90	49 37 38 44 59	
20 46 78 73 90	97 51 40 14 02	04 02 33 31 08	39 54 16 49 36	47 95 93 13 30	
64 19 58 97 79	15 06 15 93 20	01 90 10 75 06	40 78 78 89 62	02 67 74 17 33	
05 26 93 70 60	22 35 85 15 13	92 03 51 59 77	59 56 78 06 83	52 91 05 70 74	
07 97 10 88 23	09 98 42 99 64	61 71 62 99 15	06 51 29 16 93	58 05 77 09 51	
68 71 86 85 85	54 87 66 47 54	73 32 08 11 12	44 95 92 63 16	29 56 24 29 48	
26 99 61 65 53	58 37 78 80 70	42 10 50 67 42	32 17 55 85 74	94 44 67 16 94	
14 65 52 68 75	87 59 36 22 41	26 78 63 06 55	13 08 27 01 50	15 29 39 39 43	

Table A7.1.
Table of random numbers (from Table XXXIII of Fisher and Yates (1963)).