No cover image available

Critical Appraisal of Epidemiological Studies and Clinical Trials (3rd edn)

Mark Elwood

https://doi.org/10.1093/acprof:oso/9780198529552.001.0001

Published: 22 February 2007 **Online ISBN:** 9780191723865 **Print ISBN:** 9780198529552

Search in this book

CHAPTER

4 Selection of subjects for study

J. Mark Elwood

https://doi.org/10.1093/acprof:oso/9780198529552.003.04 Pages 75–122

Published: February 2007

Abstract

This chapter deals with how subjects are selected for inclusion in analytical and intervention studies. To interpret a study, one must assess the problems that may have been overcome, or introduced, by the methods used to select subjects for the study. In observational cohort studies, one has to select subjects who are or have been exposed to the putative causal factor and a suitable comparison group who are unexposed or less exposed. In the case-control design, one has to select subjects in whom the outcome has occurred (cases) to compare with subjects in whom the outcome has not occurred (controls). In intervention studies, one selects subjects who are suitable and willing to have either the intervention being assessed or the alternative with which it is compared, which may be no intervention. Part 1 of this chapter covers the general principles, and Part 2 illustrates the application of these principles to each type of study. Self-test questions are provided at the end of the chapter.

Keywords: subject selection, analytical studies, intervention studies, observational studies, putative causal factor

Subject: Public Health, Epidemiology

Collection: Oxford Scholarship Online

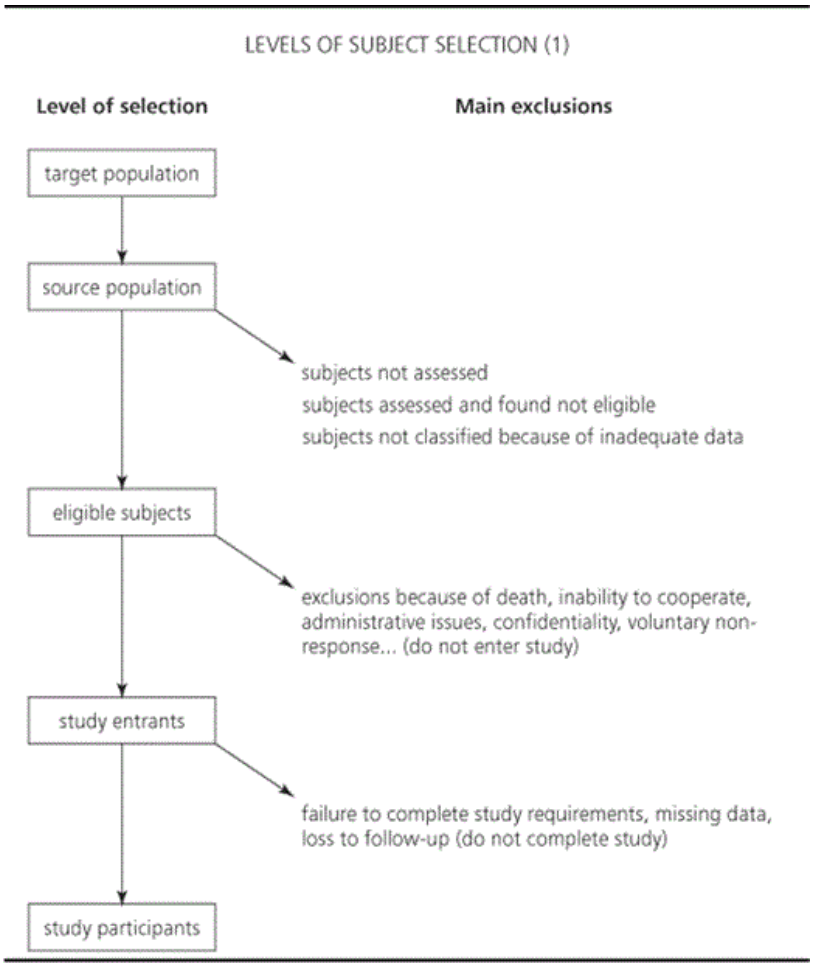
> On the 20th of May, 1747, I took twelve patients in the scurvy, on board the Salisbury at sea. Their cases were as similar as I could have them. They all in general had putrid gums, the spots and lassitude, with weakness of their knees. They lay together in one place ... and had one diet common to all... Two of these were ordered each a quart of cyder a-day. Two others took twenty-five gutts of elixir vitriol three times a day ... Two others took two spoonfuls of vinegar three times a day ... Two of the worst patients ... were put under a course of sea water ... Two others had each two oranges and one lemon given them every day ... The two remaining patients, took the bigness of a nutmeg three times a-day, of an electuary recommended by an hospital surgeon ... The consequence was, that the most sudden and visible good effects were perceived from the use of the oranges and lemons.
>
> —James Lind: A treatise of the scurvy; 1753

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025

This chapter deals with how subjects are selected for inclusion in analytical and intervention studies. To interpret a study, we must assess the problems that may have been overcome, or introduced, by the methods used to select subjects for the study. Again, we shall be dealing with the simplest possible situation, where two groups of subjects are being compared. In observational cohort studies, we have to select subjects who are or have been exposed to the putative causal factor and a suitable comparison group who are unexposed or less exposed. In the case–control design, we have to select subjects in whom the outcome has occurred (cases) to compare with subjects in whom the outcome has not occurred (controls). In intervention studies, we select subjects who are suitable and willing to have either the intervention being assessed or the alternative with which it is compared, which may be no intervention. Part 1 of this chapter will cover the general principles, and Part 2 will show the application of these principles to each type of study.

Part 1. Principles of subject selection: Target, source, eligible, entrant, and participant populations

As the derivation of subjects in a study is sometimes quite complex, we shall use five terms to describe the selection process in most studies (**Ex.** 4.1).



4.1 Formation of different levels in subject selection

p. 76 The information in any study is derived from the individuals who complete the study and contribute information to it, and we refer to these subjects as *participants*. They may be the same as the study *entrants*—those who enter the study—but some people may enter the study but not complete it. The study entrants are

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025

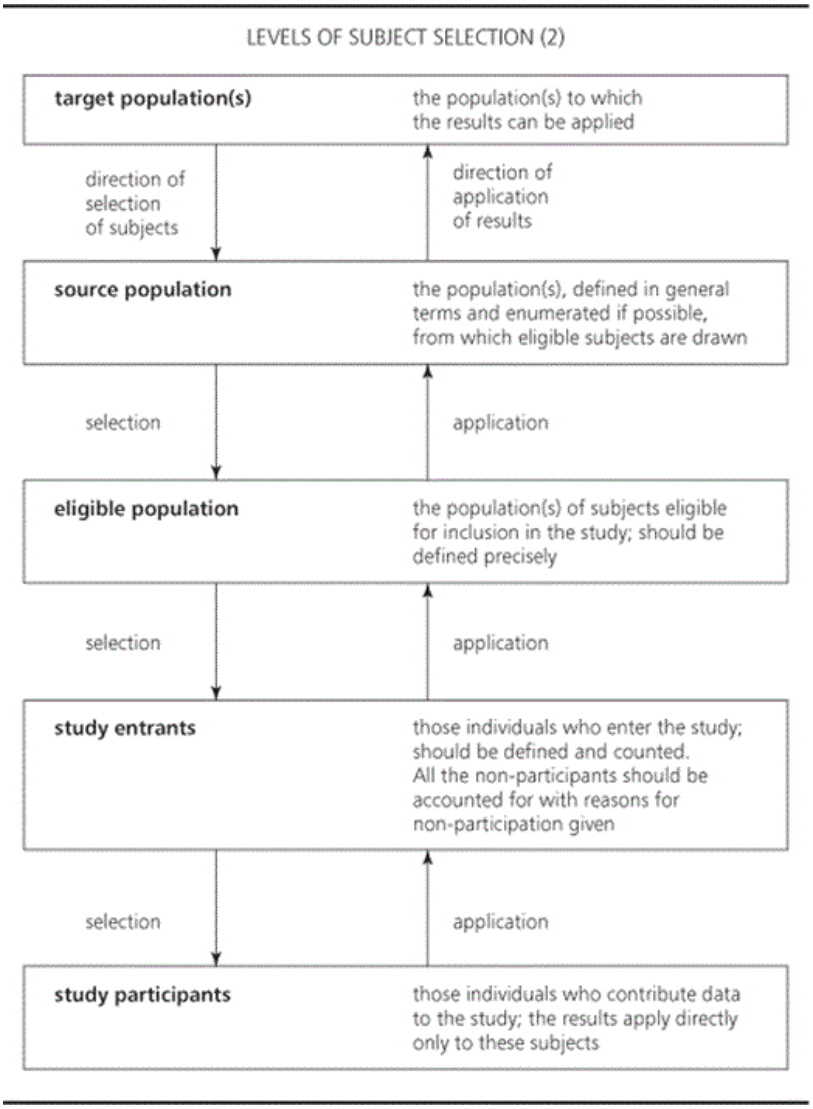
derived from the *eligible* population—those individuals who have been defined as eligible for entry into the study. Some members of the eligible population will not become study entrants. This may be because they are not invited although eligible, or they are unable to participate because of death, illness, administrative, or
p. 77 confidentiality issues, or they do not wish to participate. Such subjects do not ↳ enter the study. Eligible subjects may also fail to become study participants because although they enter the study, they do not complete its requirements and undergo the procedures or provide the data that are necessary. The correct handling of subjects who enter but who do not complete the study is an important issue, particularly in intervention trials, and will be discussed specifically in that context. Some individuals may complete part of a study (e.g. they may complete a clinical examination, but not provide a blood sample), and so they will be participants for some analyses, but only study entrants for others.

The eligible population is in turn a subset of the *source* population. The source population is determined by practical considerations, and might consist of patients in a hospital or in an individual doctor's practice, members of a particular community, a workforce, or some other group. For some studies the source population can be strictly defined and enumerated, and the proportion eligible can be calculated; in other studies the source population cannot be measured exactly, although it still needs definition. Within the source population there will be four groups of subjects: those who are eligible, those who are adequately assessed and found not to be eligible, those who cannot be classified because of inadequate information, and those who are not assessed because of lack of resources, unavailability, or other reasons.

To have practical value the study results must be applicable to subjects other than those in the original source population; for example, a study of medical treatment needs to give information that will be relevant to future patients. We shall call the population to which we aim to apply the results the *target* population, or rather target populations; unlike the other entities in the scheme, the target population is not fixed, and its definition can be modified by information from outside the study results.

In terms of subject selection, these five levels give five successively smaller subsets of subjects, each derived from the one preceding it (**Ex.** 4.2). In terms of the application of the study results, they are five successively larger populations, for each of which a further generalization of the results derived from the participants is needed.

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025



4.2 The different groups of subjects to be considered in interpreting studies

As an example, consider a clinical trial assessing different treatments in the management of acute myocardial infarction, carried out, as most such trials are, in a major teaching hospital. The study *participants* are those patients who enter the study, are randomized, and provide outcome data; their outcome information is used in the results. Some subjects may be study *entrants*, but for some reason their outcome information is not available, and so they are not full participants as they do not contribute to the key analysis. The *eligible* population consists of all patients with an appropriate diagnosis seen at the participating hospital, within
p. 78 preset limits of age and perhaps other factors, who do not ↳ have the various clinical contraindications that will be defined in the trial protocol. The study will be of little value unless we can assume that the results based on the study participants will apply to this eligible population. If many subjects enter but do not complete the trial, many others do not give their consent to enter the trial, and many are excluded for reasons other than
p. 79 those ↳ stipulated in the trial protocol, there will be substantial differences between the eligible and participant populations. Then we have to question whether the results apply to the eligible population.

A low participation rate raises questions of interpretation. For example, Slanetz *et al*. [1] assessed the opinions of doctors on whether mammograms would be better read immediately by one radiologist, or sent away for reading by two radiologists. Of 278 doctors responding to a questionnaire, 90 per cent favoured off-site double reading. This seems a conclusive result, but is based on a survey of 1000 doctors, and a response rate of 28 per cent. The results are valid only if the views of those who returned the questionnaires (the participants) are

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025

similar to those of the whole group sent questionnaires (the eligible population); as the opinions of those who did not respond are unknown, this cannot be ascertained.

For the trial of treatment for myocardial infarction, the *source* population consists of the patients admitted to the teaching hospital, or to the particular clinical unit, over a certain period of time. In principle all such patients should be assessed to see if they are eligible for the study. The *target* population is much wider, and will include patients seen in other geographical areas, perhaps even other countries, and certainly include patients seen at a future time. The definition of the target population will reflect the eligibility criteria, but also the characteristics of the source population, with regard to how individuals become part of the source population. The particular issues will be specific to the subject matter. For example, if the trial concerns therapy for myocardial infarction given immediately on admission to the clinical unit, results based on a unit which has a very rapid referral procedure from the community might not be applicable to another institution which admits only patients who have survived a considerable time since the infarct.

The choice of the source population will affect the interpretation of the results. For example, studies assessing whether children who had convulsions related to fever (febrile convulsions) also had an increased risk of further convulsions showed results ranging from 60 per cent subsequent risk down to 2–5 per cent [2]. The studies showing high risks were based on hospitals or speciality clinics, relating to children referred to them, whereas the studies with low subsequent risks were based on children identified through community-based or primary care sources. The hospital studies relate to children with more severe initial disease, which probably explains the higher subsequent rate of convulsions recorded; differences in subsequent ascertainment and follow-up methods could also be important. Similarly, the remission rate in patients with leukaemia varied from 44 to 85 per cent, depending on the eligibility criteria applied [3].

The procedure for selection of subjects who participate in the study can affect not only how widely the results of the study can be applied, but also whether the results of the study are in fact valid. To go further, it is helpful to distinguish two important aspects of study validity.

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025

The distinction between internal and external validity

All these studies involve a comparison between, at the simplest level, two groups of subjects. Thus, in the cohort study or intervention trial we compare subjects who have been exposed to the putative causative factor with subjects who have not been exposed. The *internal validity* of a study is a measure of how confident we can be that a difference in outcome between these two groups can be attributed to the effects of the exposure or intervention. The alternative explanations, which will each be discussed in detail in subsequent chapters, are that the observed difference in outcome between the groups being compared is due to *bias* in the way the observations are made, to differences between the groups in terms of other relevant factors (*confounding*), or to *chance* variation. As an example of a study with high internal validity, consider an experiment to test the cancer-causing potential of a chemical. This can be done by taking a large number of laboratory rats, bred from the same genetic strains and kept under identical conditions of diet, environment, and handling, and from these randomly selecting some animals to receive the chemical in their food, while the other animals receive a similar amount of an inert substance. The outcome would be determined by post-mortem examination of all animals at the end of their natural lifespan to determine the prevalence of cancers, with these examinations being done by a pathologist who is unaware of which animals have been given the chemical. In such a study the possibility of the observations of cancer occurrence being biased can be dismissed, the likelihood of there being some systematic difference between the animals who received the chemical and those who did not is small, and if adequate numbers are used the possibility of chance variation will be small. Therefore it is relatively easy to interpret differences in cancer occurrence between the exposed and unexposed animals as reflecting a cause and effect relationship; this ease of interpretation is due to the high *internal validity* of the study.

In contrast, consider a study attempting to look at the relationship between regular exercise and heart disease, in which a group of men who report that they take regular exercise is compared with a group of men who do not take regular exercise, with the outcome of the study being determined by the diagnoses of heart disease made by the subjects' doctors over the following few years. A difference in the recorded frequency of heart
p. 81 disease between these ↳ groups could be due to differences in ascertainment, for example if there were differences in the frequency with which subjects visit their doctors, or differences in the doctors' diagnostic criteria, or differences in their record keeping. A difference in outcome could also occur because of other differences between the two groups of men that could affect their frequency of heart disease, such as variations in cigarette smoking or diet. If the two groups of men being compared were small, the likelihood of the difference seen having arisen through chance variation might be considerable. We would say that such a study has low internal validity.

The *external validity* of a study refers to the ability to apply the results of the study to a wider population. Despite the very high internal validity of the rat experiment described above, we would hesitate to use the results to conclude that the chemical causes cancer in humans, because the species, the dosages given, the route of administration, and various other factors differ between the experimental situation and the situation which interests us. An epidemiological study of the same topic, for example comparing workers who use the chemical in their job with workers who do similar jobs but without such exposure, could give us a result that would have much higher external validity.

It is obvious that the best studies are those that have high internal validity and also high external validity; however, such studies may be difficult or impossible. Often the design considerations which help to increase the internal validity of a study may work against its external validity. Difficult choices in study design often have to be made. Going back to the example of the comparison between exercising and non-exercising men, one could argue that this study might have acceptable external validity as it is, after all, looking at the topic in free-living individuals. However, the internal validity of that study is so low that its external validity is irrelevant. It is important to realize that *external validity is useful only if the internal validity of a study is*

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025

acceptable; studies with very low internal validity have very little value. Studies that have high internal validity always have some value, even if the external validity is low. Therefore we can conclude that in designing and interpreting studies we need to pay attention to both internal validity and external validity, but internal validity is the more important.

In considering a study, each step in the chain from target population to study participants should be examined. We should ask whether the losses seen at each point produce differences between the groups of individuals being compared which may compromise the internal validity of the study, or produce a limited or atypical group of study participants which may compromise the external validity of the results.

p. 82

Selection bias and its effects

Selection bias can be regarded as any potentially misleading effect caused by the way subjects are selected to participate in a study. The selection processes can potentially affect both the internal validity and the external validity of a study, and can result in changing the hypothesis being assessed (**Ex.** 4.3).

SELECTION BIAS

Selection bias can have three types of effects on a study.
It may:

affect the internal validity

affect the external validity

modify the hypothesis being tested

4.3 Three effects of selection bias on a study

Effects on internal validity

First, selection effects can influence the *internal validity* of the study. This will occur only if the selection process has different effects on the different groups being compared. Suppose we compare the frequency of smoking in men and women by sending a questionnaire to all residents in a community. The response rate could well be higher for women than for men (which we will know if the gender distribution of the source population is known), and it may be lower for smokers than for non-smokers (we may not know that without further information). The survey will then underestimate the prevalence of smoking, and this underestimation will be greater in women. Thus the internal validity of the study in assessing differences in smoking between men and women is compromised.

Subjects who participate readily in a study often differ from those who are less enthusiastic. In an early demonstration of this, in a survey of psychosocial issues in the 1950s major psychosomatic problems were found in three out of 20 families who showed good cooperation with the survey, but in 11 out of 17 families who were less cooperative [4]. Selection factors affect internal validity only if they have *different* effects on the groups of subjects being compared within the study; this is the important distinction from the effect on external validity discussed next. As we have seen, internal validity is the more important concept, and so the primary objective in designing appropriate selection procedures is to preserve internal validity.

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025

p. 83

Effects on external validity

Secondly, selection issues can affect the *external validity* of the study. If a trial of two treatments for myocardial infarction is restricted to patients who are male, aged under 55, and have a particular pattern of infarction, the results can be directly applied to a target population which shares these features, but extension beyond needs to be justified by other evidence. Thus the selection criteria control the nature of the target population, and so limit its external validity. As these selection restrictions apply to both the groups being compared, they should not impair the internal validity. External validity may also be influenced by the participation rate, i.e. the proportion of eligible subjects who participate in the study, as a low participation rate makes it more likely that the participants are not representative of the eligible population.

Examples of studies with limited external validity which have already been given include the survey of doctors about mammography, which suffered from a very low response rate, and the studies of subsequent convulsion frequency in children who have had febrile convulsions; here, each individual study may have been internally valid, but their results depend greatly on the sources from which the subjects included were chosen.

Effects on the hypothesis being tested

The effects of selection bias may mean that the study as performed tests a different hypothesis to that originally envisaged. Consider a case–control study of the causes of 'rheumatoid arthritis'. The selection procedure used to identify cases may mean that the study actually assesses possible causative factors for 'rheumatoid arthritis that is sufficiently severe to lead to hospital treatment', which may be a hypothesis considerably different from that originally envisaged.

This issue is closely related to that of misclassification. External validity will be highest where the cases in the case–control study, or the exposed group in a cohort study, can be regarded as representative of all cases or of all exposed individuals in the source population. However, frequently the attempt to maintain high external validity introduces the risk of inaccuracies in the definition of these study groups, so that they include non-cases or non-exposed individuals. For example, in the case–control study of rheumatoid arthritis there are choices between the two extremes of entering all individuals in a defined community (the source population) who have any type of diagnosis of rheumatoid arthritis, or of entering only those who have rheumatoid arthritis defined by specific criteria and supported by specific laboratory and radiological investigations. The latter procedure will lead to less misclassification, but if a full investigation is performed only on patients with
p. 84 severe ↳ disease, the participants will be less likely to be representative of all individuals with rheumatoid arthritis. The balance between these two options will depend on the particular circumstances of the investigation.

For example, several case–control studies in the 1970s (e.g. [5]) compared women with endometrial (uterus) cancer with control groups selected on various criteria, and showed a much higher frequency of the use of oestrogens, prescribed mainly for the control of menopausal symptoms, in the endometrial cancer cases. However, a subsequent case–control study showed no association; the controls for that study were chosen as women who had been investigated for endometrial cancer, but found not to have it. The argument made for the comparison was that using such a control group would ensure that no controls had unrecognized endometrial cancer, i.e. misclassification was avoided. However, the comparison being made was between women with endometrial cancer and other women investigated because of similar symptoms, such as bleeding. If oestrogens cause both endometrial cancer and other conditions that lead to bleeding, the lack of association reported in that comparison would be expected, and would not assess whether the risk of cancer was increased. A further study compared women with endometrial cancer with three other groups: women having investigation for gynaecological symptoms, other gynaecological patients, and a community-based group. Oestrogen use was highest in the women with cancer and the first of these comparison groups, showing that it

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025

caused both endometrial cancer and other non-cancer conditions leading to similar investigations [6]. One of the first of these studies is discussed in detail in Chapter 14.

There have been several investigations of the relationship between psychological parameters, previous life events, and breast cancer by studying women attending breast clinics for diagnosis. This has the advantage of allowing interviews or questionnaires to be applied prior to diagnosis, avoiding the response bias that might arise after the diagnosis. Factors that are more common in women who are subsequently diagnosed with breast cancer than in women without cancer have then been interpreted as causal factors for breast cancer. However, the comparison being made is not between women with breast cancer and women representative of the general population, but between women with breast cancer and women with other breast problems which bring them to a diagnostic clinic. A positive association with breast cancer could arise because the factor prevents other breast conditions that would lead to attendance at that clinic; one such factor is oral contraceptive use, which decreases the risk of some benign breast conditions.

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025

p. 85

Methods of reducing selection biases

Selection effects are of two kinds: first, effects of incomplete *participation*; secondly, effects produced by the selection *criteria*. The investigator should have knowledge and control over the selection criteria; only partial control of participation is possible.

Reporting, and optimizing, participation and response

The *participation rate* indicates how the participants in a study may differ from the eligible population. The participation rate is defined as the number of study participants divided by the number of eligible subjects. It is particularly useful to compare the participation rates of the different groups of subjects in the study, as differences may affect the internal validity of the study.

The *response rate* is one component of the participation rate. The response rate is the number of study participants divided by the number of eligible subjects who were identified, contacted, and asked to participate; it is a measure of the completeness of voluntary response by the subjects. As such, it is useful and indicates one important part of the selection process. As it does not account for losses by mortality, failure to locate, exclusion by doctors, and so on, it should not be used as the only or main estimate of participation, although it frequently is in publications, perhaps because it is often impressively high. The participation rate is, of course, always lower than, or at the maximum equal to, the response rate.

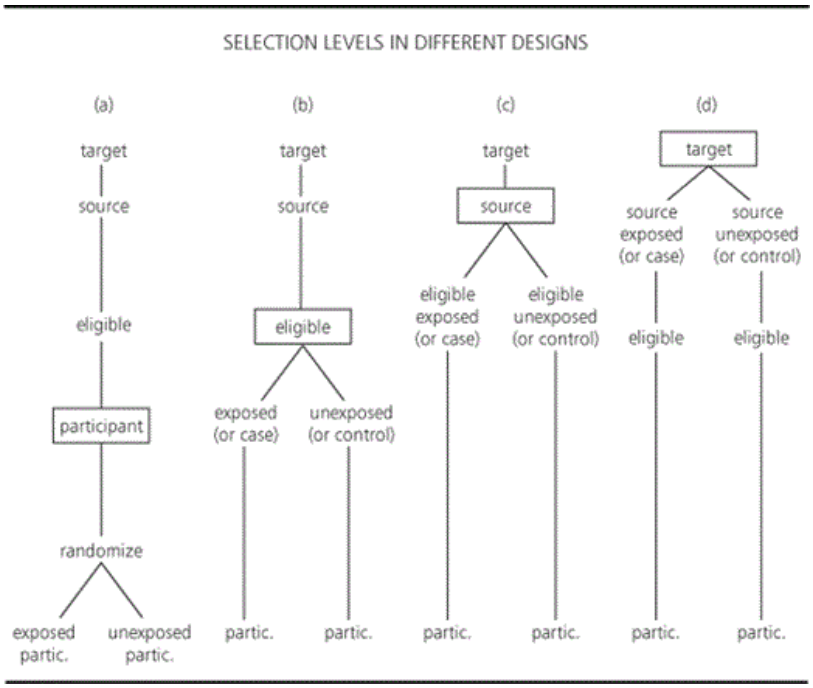
Ideally, all studies should report on the participation and response rates, but this is not always done. In a survey of the published information on participation in 355 epidemiological studies published in 10 major journals in 2003, information on participation was given in only 32 per cent of cohort studies, 44 per cent of case–control studies, and 59 per cent of surveys. The information given was sparse; for example, participation and subject response rates could both be calculated in only 16 per cent of reported case–control studies [7]. Participation rates declined over the period 1970 to 2003.

Obtaining high response rates in surveys and questionnaires is a major subject in its own right. A systematic review of randomized controlled trials of strategies to increase response rates to postal questionnaires based on information up to 2003 included 372 trials [8]. The review reported that many methods had been shown to increase response rates. These included monetary and non-monetary incentives, a promotional message on the envelope, a more interesting topic, pre-notification of the subjects, follow-up contact, shorter
p. 86 questionnaires, providing a second copy of the questionnaire at follow-up, ↳ mentioning an obligation to respond, university sponsorship, personalized questionnaires, coloured as opposed to black or blue ink,

stamped returned envelopes instead of franked returned envelopes, an assurance of confidentiality, and first class mailing. The response rate was reduced when sensitive questions were included, when questionnaires began with the most general questions, and when participants were offered the opportunity to opt out of the study. Of course the various trials reviewed were all in different contexts, and not all factors would be relevant to all questionnaires.

Selection effects with different study designs

As the maintenance of internal validity is the most important objective in study design, the stronger study designs are those in which the selection criteria apply equally and with the same effects in each of the groups being compared. The outline diagram of a randomized intervention trial (**Ex.** 4.4, type (a)) illustrates the value of this design. Only subjects who are eligible and have given their consent to the study, including consent to randomization and to each of the interventions being offered, enter the study. The selection criteria are identical to the point of randomization. The factors influencing participation act prior to randomization, and so will affect the intervention and comparison groups equally. From the point of randomization, all subjects will be included in the analysis, irrespective of whether they accept the prescribed intervention and complete the follow-up procedures or not (this issue will be described more fully when we consider the role of randomization in preventing confounding in Chapter 6). As the selection criteria apply to each of the groups in an identical fashion, selection issues will not affect internal validity. However, the external validity of this design may be quite limited, as the strict eligibility criteria and the requirement for consent prior to randomization may make the participant group a relatively small and perhaps unrepresentative sample of the eligible and source populations.



4.4 Study designs showing different selection schemes. From (a) to (d), the pathways for selection of groups to be compared become more different, and so the possible influence of selection on *internal* validity increases. (a) is a randomized trial design; (b), (c), and (d) are varieties of non-randomized intervention, cohort, or case–control designs. Partic. = participants

The effects of selection on internal validity become more severe as the design departs from the ideal of the randomized trial. Exhibit 4.4, types (b), (c), and (d), shows designs in which the differences in selection appear at the levels of the participant, eligible, and source populations respectively. As an illustration, consider the

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025

design of a prospective cohort study comparing women using oral contraceptives with those using other methods of contraception in terms of later disease. The ideal scientific design would be a randomized trial, but clearly this is ethically impossible. The next strongest design is one in which a suitable source population is identified and eligibility criteria are set which are identical for exposed and unexposed women; this is design
p. 87 (b). For example, the eligible population could be defined as all women ↳ attending a defined group of doctors who start a new contraceptive method (oral contraception or other method); an example of this design will be described subsequently. By having the same eligibility criteria for both groups, some similarity is ensured; however, the eligible groups of oral contraceptive users and non-users may differ in other factors that affect their outcome rates, and the participation rates for users and non-users may differ. The analysis of the study needs to take account of these possible differences between the groups.

If the eligibility criteria for oral contraceptive users and for the comparison subjects are not the same, greater differences will be introduced, and this becomes design (c). For example, oral contraceptive users might enter the study from the time of their first use of an oral contraceptive, but it may be convenient to enrol comparison
p. 88 subjects using other contraceptive methods ↳ whether they were just starting on these methods or had used them for some time. This difference in eligibility criteria could introduce further differences between the groups being compared, giving greater effects on internal validity.

Another design would be of type (d), where the source populations are different. For example, the oral contraceptive users might be identified as women who had received their contraceptive prescriptions from a certain clinic, while comparison subjects might be taken as women using other methods of contraception, identified in other ways. Therefore the source populations are different, and factors affecting this difference, such as factors influencing whether women go to a particular clinic, can then contribute to the differences between the exposed and unexposed groups.

The same types of consideration apply to case–control studies. Therefore it is helpful in assessing or designing studies to define the participant, eligible, source, and target populations, as this may illustrate where problems of validity may arise.

Selection of subjects for comparative studies

A few clear principles apply to the selection of subjects for all comparative studies, whether intervention trials, cohort studies, or case–control studies. While the details are specific to each study design, the principles apply to all these studies.

In comparative studies, there are two groups of subjects: the group of prime interest (the cases in a case–control study, the exposed group in a cohort study, and the intervention group in a trial) and the comparison group. There are four principles relating to the selection of the group of prime interest, shown in **Ex.** 4.5.

1. The groups should be what they are designed to be, i.e. the group of prime interest should truly be cases,
p. 89 or exposed, or an intervention group. If we ↳ include amongst a cohort defined as exposed some subjects who are not exposed, we will underestimate the true size of the association between exposure and outcome. Misclassification by exposure status in a cohort study, or case status in a case–control study, will bias the results of the study towards the null hypothesis. The direction of this effect is useful to note. In assessing published work, misclassification is not a serious issue in studies that show a strong association, as a reduction of the misclassification will actually strengthen the observed association. However, in interpreting studies that show no association, perhaps a true association exists and there has been sufficient misclassification to disguise it. In some circumstances quantitative estimates of the degree of misclassification can be made, and the results adjusted accordingly; this will be discussed in Chapter 5.

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025

2. The group of prime interest should be ascertained from the beginning of the factor's operation, i.e. a case group should be selected from newly incident cases, and an exposed or intervention group from the beginning of the exposure or intervention. Consider a study to look at the frequency of muscular pain in workers doing repetitive jobs in a factory. The simplest design is to go to the factory, examine workers who are doing the particular job, and find out how many of them have evidence of muscular problems. This will almost certainly underestimate the problem, as the study includes only workers who have started the job and continued it for various periods until the time of the investigation. If, instead, we study all workers who start on the job, we might find that many of them develop muscular problems and then change their job or leave the workforce entirely.

3. The exposed or case subjects should be representative of a defined eligible population. As we have seen, this defined eligible population is the essential link between the exposed or case group in the study and the comparison group.

4. The subjects must be chosen so that the necessary other investigations can be carried out, i.e. the assessment of outcome in a cohort study, exposure in a case–control design, and related factors in either. These other investigations should be carried out in a similar manner in the control groups chosen, and with similar completeness. For example, one of the main cohort studies of the effects of smoking was the study of British doctors started in the 1950s. The decision to base the study on doctors was made largely because they would be interested in participating in the study, and as they had to re-register formally each year to maintain their license to practice, the difficulties of keeping them under follow-up were minimized.

QUALITIES OF THE 'EXPOSED' GROUP IN A COHORT STUDY AND THE 'CASE' GROUP IN A CASE–CONTROL STUDY

1. should be truly 'exposed' or a 'case'
2. should be newly exposed, or be a newly incident case
3. should be representative of a defined eligible population
4. should be available for study so that necessary information can be collected in the same way as it is for the comparison subjects

4.5 Criteria for the groups of prime interest in a comparative study. Total fulfilment of all criteria is rarely possible.

p. 90 A little consideration will show that feature 4 above relates primarily to dealing with bias in observations. Feature 3 is mainly concerned with dealing with differences between the groups in regard to other factors, i.e. confounding. Feature 2 will also relate to this, and feature 1 relates to the points we have just reviewed, that the selection of subjects may affect both the internal and external validity of the study, and may modify the hypothesis under test. Good study design requires a balance between these four features; often a strategy to improve one of these features may compromise another.

Selection of the comparison subjects

The essential characteristics of the comparison group, whether it be the unexposed group in a cohort study or the unaffected group in a case–control study, follow logically from, and are equivalent to, the criteria for the exposed or case groups (**Ex.** 4.6).

1. The controls must be representative of the group they are designed to be. In cohort studies, that means they should be unexposed or minimally exposed. In many case–control designs, controls are defined as free of the outcome of interest, and so should be selected to fit that definition. Alternatively, the design

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025

may call for the controls to be a sample of the whole population, as discussed in Chapter 3. As pointed out above, misclassification in this regard will have the effect of biasing the measured association towards the null value, and cannot exaggerate the true association. Therefore a small degree of misclassification may be acceptable, particularly if to avoid it would compromise other valuable parts of the research design.

2. The control group should be chosen so that the relevant information can be collected in a manner analogous to that used for the exposed or case series.

p. 91

3. ↳ A useful general concept is that comparison subjects should be representative of the unaffected (or of all) members of the same eligible population that provided the exposed or case subjects. While this general statement is usually applicable, the choice of appropriate comparison subjects is a complex issue and the ideal characteristics cannot be fully described in a simple inclusive statement. Therefore we shall explore this issue more fully in the context of the different study designs.

QUALITIES OF THE COMPARISON GROUP IN A COHORT OR CASE–CONTROL STUDY
1. Should be truly 'non-exposed' or 'non-diseased'
2. Should be available for the study so that necessary information can be obtained in the same way as it is for the exposed or case subjects
3. Should be representative of a defined eligible population, analogous to that for the 'exposed' or 'case series'; this can be modified in that comparison subjects may be chosen specifically to be similar to the exposed or case series with regard to particular factors

4.6 Criteria for the comparison groups in a cohort or case–control study. Total fulfilment of all criteria is rarely possible.

Assessment of selection issues in a completed study

We have seen how the selection of subjects for a study can influence both internal and external validity, and affect the hypothesis under test. In assessing a study, it is useful to examine the component populations, and a simple scheme is shown in **Ex.** 4.7. The general questions, applying to the whole study, are relevant to external validity and to the hypothesis; the questions concerning differences between the groups being compared are also relevant to internal validity.

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025

ISSUES IN THE SELECTION OF SUBJECTS	
General questions	Comparison of the groups
What is the definition of the eligible population?	Are the definitions comparable?
How do the participants relate to the eligible population?	
What is the participation rate?	Compare the groups
What are the reasons for losses, and how frequent is each?	Compare the groups. Are differences likely to affect internal validity? Do the losses compromise external validity?
What is the source population?	Is it the same for each group?
How do the source and eligible populations relate?	Is the relationship similar for each group? Is the main result likely to apply to the source population?
What is the target population?	Is the main result applicable to this target population?

4.7 Selection of subjects. An outline scheme to assist in the consideration of issues of subject selection in a particular study. The questions should be considered generally, and specifically in regard to the comparability of the relevant groups: exposed and unexposed in cohort and intervention studies; affected and unaffected in case–control studies.

p. 92

Part 2. Application of the principles of selection of subjects to each type of study, and examples: Selection of subjects in randomized intervention trials; intention to treat analysis

This design is conceptually the simplest. In the design used most commonly, exemplified by the trial of treatment for tuberculosis discussed in Chapter 1 (Ex. 1.8), a group of participants is selected on the basis of eligibility criteria and informed consent, and divided into the intervention and comparison groups by randomization. In an alternative design, eligible subjects are identified and randomization is then performed; those randomized to be offered the intervention are then approached for consent and participation, while those randomized to the comparison group receive their normal care, and may indeed be unaware that the trial is proceeding. This design is often used in large-scale interventions comparing a new intervention with routine care, as consent is required only for those who will be offered the new intervention. Thus, in a trial of breast cancer screening, some 64 000 women were randomized into two equal groups; one group was offered screening and their consent to screening was sought, while the other group continued with normal care without any need for consent [9].

We can apply the four principles of subject selection, which were shown in Ex. 4.5, to the intervention group. Ideally, those randomized to the intervention receive the intervention, and the comparison group do not. In practice, this is rarely likely to happen without some compromises. Some subjects randomized to the intervention may never receive it because they do not accept it, or clinical contraindications arise, or there are administrative difficulties, and some of those who start may not continue it for very long. Therefore there some misclassification is produced which will reduce the difference between the intervention and comparison groups. Despite this, the appropriate analysis compares the ultimate outcome in the original total groups defined by the randomization. Only this comparison maintains the advantages of the randomized design; this is the *intention to treat* analysis. If comparisons are based only on the subjects who accept the intervention, or complete it, the comparisons are open to all the difficulties of comparing non-randomized cohorts. Thus in the breast cancer trial noted earlier, death rates were compared between all women in the control group and all

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025

women in the group randomized to be offered screening, although only about two-thirds of them accepted the offer.

Similarly, the comparison group may be influenced by the intervention. This may mean that the association actually assessed is different from that originally envisaged. For example, in a trial of health education, an
p. 93 intervention ↳ group may be selected by randomization and offered a new education programme, but the comparison group, which is not offered the intervention, may make similar changes themselves. This is referred to as *contamination* or *dilution*. The comparison actually being made is between the specific intervention and the other changes affecting the control group. In the Multiple Risk Factor Intervention Trial (MRFIT) in the USA 12 866 men at high risk of coronary heart disease were randomly allocated to a special intervention programme or to normal care [10]. Men allocated to the intervention programme showed substantial reductions in blood pressure, cholesterol levels, and smoking. However, substantial, although lesser, reductions were also seen in the group randomized to normal care. The trial results showed only a small and non-significant reduction in mortality from heart disease in the intervention group. This was an '*open*' trial, i.e. both the participants and their own doctors were aware of the trial. For men in the control group, the information from the assessments made was sent to their own doctors, although without any recommendations for action. The changes seen in the normal care group could have been produced by several factors: enrolment in the trial, even although they were randomized to the normal care group; those volunteering for the trial being already motivated to change; the impact of the new knowledge of risk factors from the examinations carried out in the trial; or the doctors of men in the normal care group making their own intervention recommendations. Carrying out the trial in a more rigorous fashion, without giving feedback on the results of the regular examinations to men in the normal care group, was regarded as unethical. Similarly, in trials offering screening mammography to randomized groups of women, substantial numbers of women in the non-intervention group will also receive mammography through their own initiative or through other programmes [11].

As the interventions are under the control of the investigators, the issue of being newly exposed should be well defined. Sometimes prior exposure to a similar intervention (e.g. the same drug) is expected to influence the effect of the exposure under test, and so lack of prior exposure may be used as an eligibility criterion.

The great strength of the randomized design is with regard to representativeness, i.e. both the intervention and comparison groups are representative of a defined eligible or participant population. The random selection procedure, if done on adequate numbers, will result in two groups which are likely to be similar in terms of any particular factor. Further, the exposure is added independently to one group, and so should not be associated with other factors influencing the outcome. Thus it is reasonable to assume that the frequency of outcome seen
p. 94 in the comparison group would also be seen in the exposed ↳ group if the exposure had not occurred or had no effect. This does depend on adequate numbers, so, a small randomized trial may well show differences between the groups in relevant factors, as will be discussed in Chapter 6.

Also, in a randomized design the methods of ascertaining outcome and other relevant factors can be identical in the two groups. There is the special opportunity, rarely provided in the other designs, to use single-blind and double-blind techniques, i.e. study designs where the subjects are not aware of which intervention they are receiving (*single-blind*) or where neither the subjects nor those making the observations of outcome are aware of this (*double-blind*). This is obviously easiest with drug trials, and more difficult with trials of other interventions. (In a *triple-blind* design, in addition those analysing the data only have a code indicating which intervention has been given, and do not know what the code means).

Randomized trials have a potential weakness in that the application of strict eligibility criteria and management protocols may mean that the subjects who participate in the trial are unrepresentative of the wider group of relevant subjects in the community. This raises the question of whether the results of the trial can easily be applied to a wider population. Another criticism of trials is that many are designed to address

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025

scientific issues predominantly, and there is a need for more clinical trials to address questions faced by decision-makers providing health care. Such trials have been called *pragmatic* or *practical* clinical trials [12]. They are designed to address important clinical or management issues, to include a diverse population of participants recruited from a range of health care settings, and to collect data on a wider range of health outcomes, adding patient-centred outcomes, such as symptom control and quality-of-life measurements, and economic outcomes to the more traditional clinical outcomes.

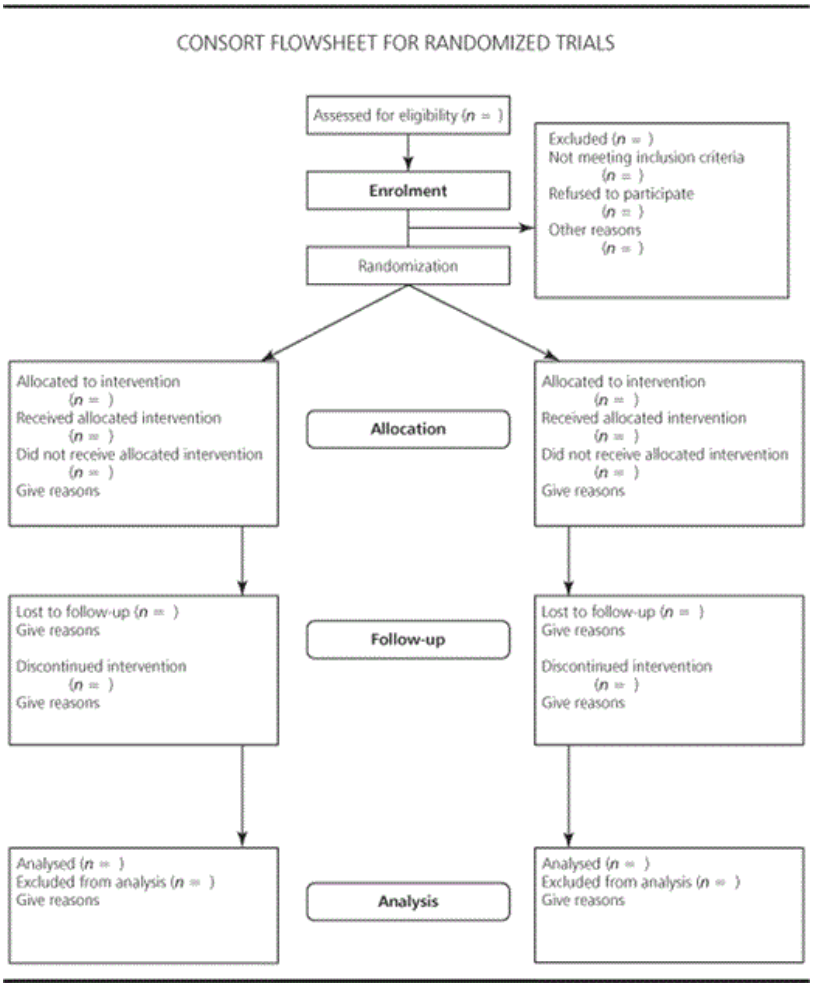
In the design of clinical trials, the number of subjects required is estimated by methods that are discussed further in Chapter 7. In many trials, particularly large and long-term trials, the results are monitored as they come in by an independent data monitoring committee which is separated from the principal investigators of the trial, and can operate in a 'triple-blind' fashion. This is to fulfil the ethical ideal of stopping the trial as early as possible so that a conclusion can be reached and the optimum treatment can be offered to all participants in the trial and to other people. This issue is also discussed further in Chapter 7. The decision of whether or not to stop a trial can be complex and is often controversial. Stopping a trial too late delays results and will mean that more participants of the trial are given the inferior treatment; but stopping a trial too early can make interpretation difficult and may lose the scientific value of the trial not only for all its participants but for people in general. For example, a randomized trial was started in 1997 to investigate whether hormone
p. 95 replacement therapy for menopausal symptoms was safe in women ↳ with a previously treated breast cancer. The trial had a data monitoring committee and a preset protocol for examining the results regularly. In September 2003, after a median follow-up of 2.1 years, 26 of 174 women in the hormone replacement therapy group and seven of the 171 women in the comparison group had signs of breast cancer progression, and the trial was stopped [13]. Other trials are stopped because there is clear evidence of benefit before the predetermined study size is reached. This occurred in a trial of the benefits of folic acid in preventing birth defects, which is described in Chapter 11.

In summary, in the randomized intervention design, the prime advantages are that both groups are drawn from the same eligible or participant populations, and that the randomization is likely to lead to comparability in regard to other factors, and to similarity in how the observations of outcome are made. These characteristics, features 2, 3, and 4 on the generic list in Ex. 4.5, are given prominence over feature 1, so that the analysis by comparing randomized groups may accept a degree of misclassification because of incomplete participation of those randomized, or other influences.

The CONSORT report format for randomized trials

The CONSORT (Consolidated Standards of Reporting Trials) statement is a recommended way of describing the results of randomized trials which has been adopted by many leading medical journals [14,15]. It follows the principles set out already, and emphasizes the need to document all exclusions or departures from the protocol after randomization. The use of the format has been shown to improve reporting [16]. The CONSORT scheme is shown in **Ex.** 4.8, and is described at http://www.consort-statement.org where amendments and new publications are given. As can be seen, it deals with the trial from randomization onwards, and so addresses issues of internal validity; it does not deal with questions of external validity. Another emphasis has been on ensuring that all trials that are performed are reported, so that all the results can be assessed to avoid publication bias; this is discussed further in Chapter 8. Investigators are encouraged to register their trial and submit the protocol when the trial is designed. Many leading medical journals will now only publish the results of trials which have been registered in advance, and which are reported according to the CONSORT criteria.

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025



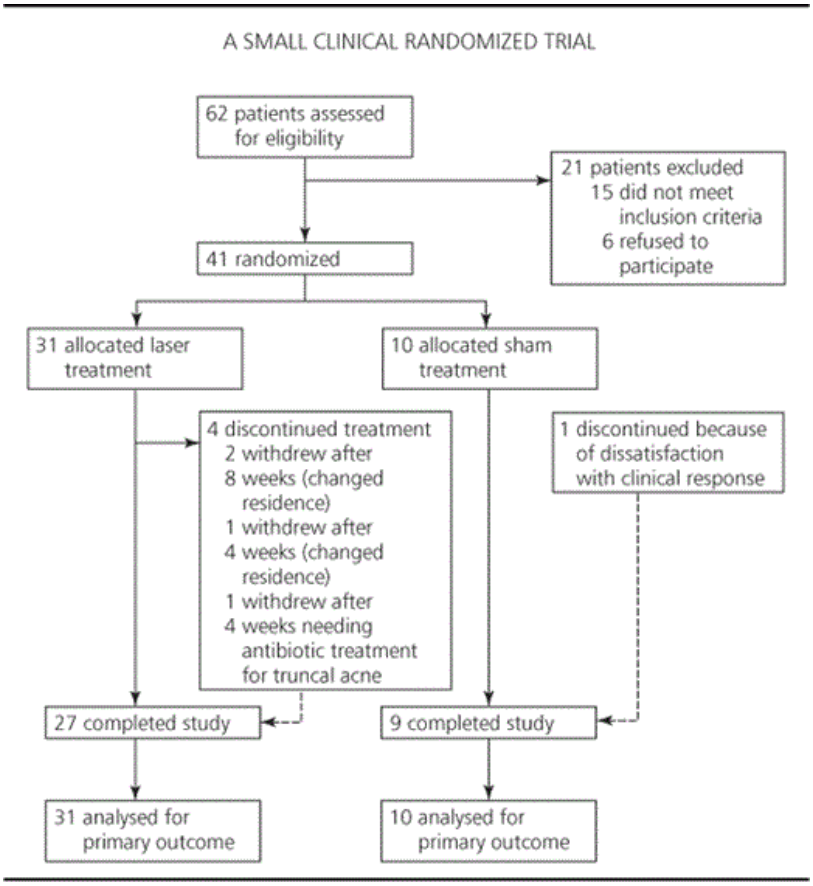
4.8 The CONSORT (Consolidated Standards of Reporting Trials) recommended flowchart to describe a randomized trial. From www.consort-statement.org [14]

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025

Examples of randomized trials

A trial of an innovative laser treatment for facial acne illustrates the issues in the selection of subjects for a small randomized clinical trial [17]. **Exhibit** 4.9 shows the conduct of the trial, following the CONSORT scheme.
p. 96 Eligible patients were selected from one hospital clinic or recruited by a public ↳ advertisement, and had to be aged between 18 and 45 years and have mild to moderate facial inflammatory acne, which was assessed on a scoring system accepted by dermatologists. Because previous treatments could affect the trial, minimum time intervals were set; for example, patients had to have had no oral antibiotic treatment during the previous 4 weeks, but no treatment by isotretinoin (a vitamin A derivative), which has a much longer effect, during the previous 52 weeks. Sixty-two such patients were assessed for eligibility, of whom 15 did not meet the entry criteria and a further six chose not to participate, giving 41 patients who were randomized. A three-to-one
p. 97 randomization ↳ scheme was used, resulting in 31 patients allocated the new laser treatment and 10 allocated sham treatment, where the disconnected laser handpiece was moved across the face in an identical manner to that of the active treatment group. All patients wore opaque goggles to protect their eyes, which also ensured that they were unaware of which therapy they received. Four of the 31 patients allocated laser treatment discontinued it for reasons which are given in the report, and one of the 10 patients allocated sham treatment discontinued it, giving 27 and nine, respectively, who completed the study. However, follow-up examinations
p. 98 were done on all patients initially randomized, and so ↳ the comparison for the primary outcome, based on a clinical assessment of the acne after 12 weeks, was made using all 41 originally randomized subjects. There was a substantial improvement in acne in the laser group, with no change in the sham treatment group. Some other secondary endpoints could only be assessed on those who completed the study, which is less satisfactory, although here the number of subjects discontinuing the trial is small.

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025



4.9 CONSORT flowchart for a randomized trial of laser treatment compared with sham treatment for inflammatory acne. Reprinted from The Lancet 362(9393), Seaton ED, Charakida A, Mouser PE, Grace I, Clement RM, Chu AC. Pulsed-dye laser treatment for inflammatory acne vulgaris: randomised controlled trial. 1347–1352. Copyright 2003, with permission from Elsevier. [17]

Although this randomized trial is small, information is given on the comparability of the laser and sham treatment groups with regard to demographic factors, previous treatments, and characteristics of the initial acne. The internal validity of the results is good. The subjects to whom the trial was offered are defined in a clinically relevant way in terms of an accepted severity scoring system, and so the results should be applicable to similarly defined patients elsewhere. However, there is no guarantee that the patients seen at this particular clinic are representative of any wider group, and verification of that would logically come from repetition of the study with other populations.

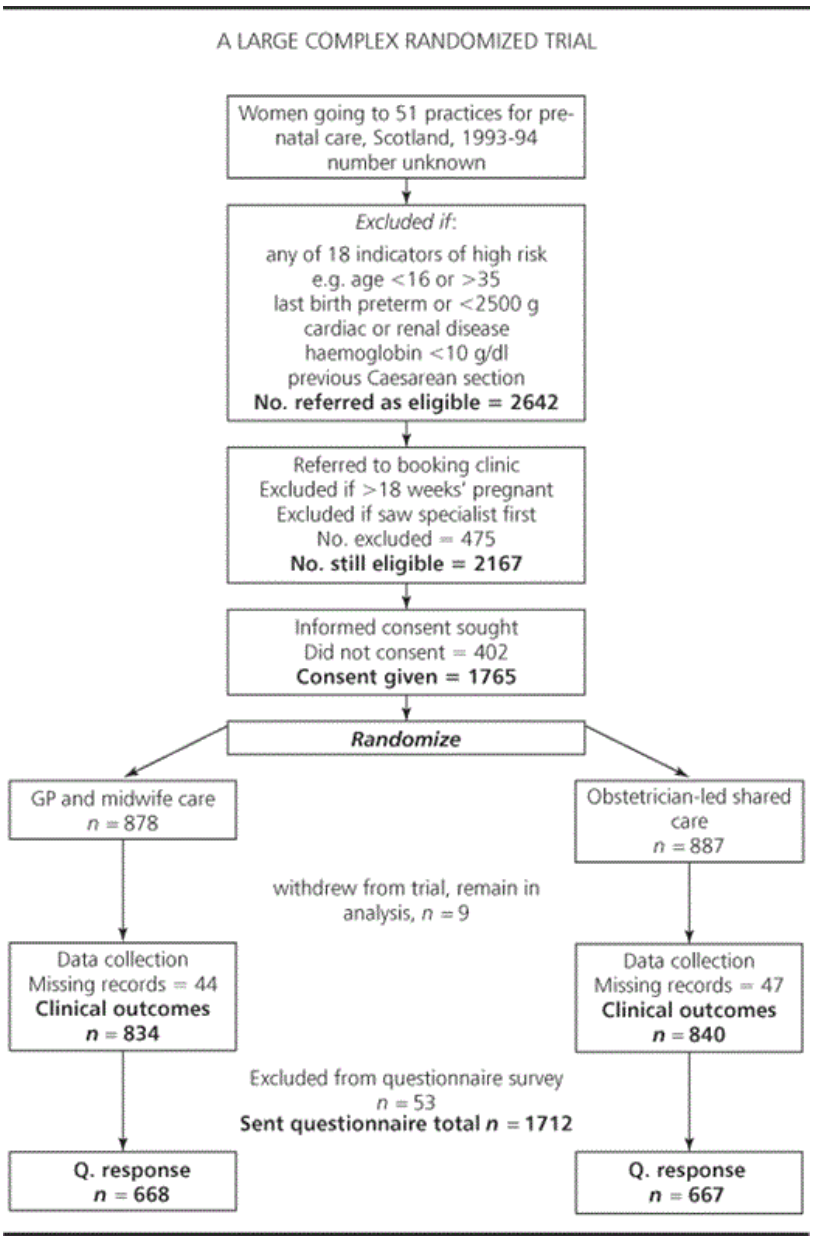
An example of a much more complex community-based trial is one comparing two methods of providing maternity care in Scotland [18]. The objective was to study women coming for prenatal care, exclude high-risk women, and assess for the other women whether a specialist-led care system (which was the norm at that time) produced any different results from a new system of care led by general practitioners (GPs) and midwives. As shown in **Ex.** 4.10, this required a great deal of set-up work, starting with obtaining the consent of 51 general practices to participate in the study so that women could be assessed and excluded if there were indicators of high risk, and obtaining the support of all the specialist obstetricians in the area. Of the 2642 women assessed as eligible, 475 were excluded because they were already beyond 18 weeks of pregnancy or had already seen an obstetrics specialist. A further 402 women did not consent to the study, leaving 1765 who were randomized into two approximately equal groups. A small number of women withdrew from the trial but could be included in the analysis. The analysis was based largely on clinical records, and further exclusions were because records were missing; to assess the patients' opinions of their care, a questionnaire was used which had an incomplete response so that the results for that part of the trial were based on smaller numbers. This

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025

was a complex design, but there is little doubt that the randomized trial design ensured that the women treated with each of the alternative systems of care were generally comparable, and the general design of the study ensured that the results are relevant to the community and health care system in which the study was done. The study showed that the new GP and midwife care system produced better continuity of care, fewer antenatal
p. 99 hospital admissions, a modest reduction in routine clinic visits, and ↳

p. 100 ↳

a lower frequency of some of the more common complications of pregnancy. The levels of satisfaction of both groups of women with their care were very similar, and high. The authors concluded that GP and midwife care was an acceptable alternative to obstetrician-led care, which was the norm at the time.



4.10 Derivation of groups in a randomized trial comparing two systems of prenatal care for low-risk women in Scotland. Data from Tucker *et al*. [18]

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025

Selection of the subjects for a cohort study

In an observational cohort study, the groups compared are the exposed group and a comparison group. The 'exposed' group should truly be exposed, and misclassification by exposure status in a cohort study will bias the results of the study towards the null value. However, misclassification is often severe in cohort studies because an indirect indicator of exposure is used. To assess the health effects of exposure to asbestos, for example, an 'exposed' group of subjects who have worked in an environment where asbestos was used may be chosen, even though many of them may have had little or no exposure. The results will demonstrate the health effects of the average exposure of this group; a real effect may be missed if there are many individuals in the group who have not had exposure.

A related issue is that the exposure may change over time. In the study of asbestos, the initial cohorts of exposed and unexposed workers may be set, but as the study follow-up proceeds some exposed workers will cease exposure and some unexposed workers will begin exposure, introducing further misclassification. If such variations are large, special types of analysis will be necessary, such as defining each subject on the basis of total length of exposure. This method was used in a study of oral contraceptive users discussed later in this chapter.

Subjects should ideally be newly exposed to the causative agent under study. As noted above, a prevalence survey of currently employed workers will underestimate a problem that leads some workers to quit. This consideration of being newly exposed must be taken along with the definition of the exposure. In studying the frequency of cancer in workers exposed to a particular chemical through their occupation, we might assume that a substantial amount of exposure would be necessary before any detectable increase would occur, and therefore we might define exposure as a minimum of 5 years occupational exposure to the chemical. In this situation, being newly exposed means that this 5-year period has just been completed, and the study may include the total follow-up period subsequent to that time for each subject, but exclude subjects who leave the workforce or change jobs before they have 5 years exposure.

Similarly, the comparison group of subjects who are regarded as unexposed, should actually be unexposed. As pointed out above, misclassification in this regard will have the effect of biasing the measured association
p. 101 towards the null ↳ value, and cannot exaggerate a true association. Therefore a modest degree of misclassification may be acceptable, particularly if to avoid it would compromise other valuable parts of the research design.

The exposed and unexposed subjects must be chosen so that the appropriate investigations can be carried out, i.e. the assessment of outcome and of related factors. These investigations should be carried out in a similar manner in the control groups chosen, and with similar completeness.

Choices with regard to the control group in cohort studies

It is helpful to concentrate on the *purpose* of the control group. In a cohort study, we measure the frequency of the outcome in exposed subjects. The function of the control group is to estimate what that rate would be in those same subjects, had they not been exposed. The frequency observed in the exposed group will depend on the effects of the exposure factor, but also on the other characteristics of that exposed group that influence the outcome. Therefore an appropriate comparison is a group of subjects who share, as far as possible, all the other factors which influence the outcome, apart from the exposure.

The best way to achieve this is by a randomized trial. However, random allocation and an intervention design are often not possible. In the randomized trial design, the control group has two properties; it is a representative sample of the original eligible population, and it is likely to be similar to the exposed group with

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025

regard to other relevant factors. Procedures for selecting controls in non-randomized studies can be logically determined by starting from one or other of these properties.

The control series can be chosen as *a representative sample of the unexposed* members of the eligible population from which the exposed subjects are also drawn—design (b) in Ex. 4.4. A useful practical guiding point is that all potential controls, if they were exposed, should be eligible for inclusion in the exposed group. This approach ensures comparability of the exposed and unexposed groups with regard to characteristics that define the eligible population.

In an observational study the subjects themselves or, in the case of therapy, their medical advisers, have chosen whether they are to be exposed or unexposed to the factor in question. This self-selection will usually mean that the exposed and unexposed groups differ with regard to other factors that influence the outcome. For example, smokers and non-smokers differ with regard to other aspects of lifestyle such as alcohol use and diet; a non-randomized comparison of patients who have been given different treatments will often be made difficult because the patient's clinical findings and current prognosis will influence the treatment given.

p. 102 If we know a great deal about the factors that influence the outcome under study, we could choose a comparison group that is *deliberately made similar to the exposed group* in terms of the other factors that determine outcome. This results in a matched design, in which for each exposed subject, one or more unexposed subjects are chosen because they share the other characteristics that affect the frequency of the outcome variable. Thus in a study of the long-term outcomes of amniocentesis, Baird *et al*. [19] identified a cohort of 1296 liveborn infants whose mothers had had amniocentesis during that pregnancy, and compared them in terms of later disabilities with 3704 control liveborns whose mothers were matched for sex, maternal age, area of residence, and time of birth. No differences were found except for an increase in haemolytic disease due to iso-immunization. Matching can give a powerful design, but it has several disadvantages. It is not often that we know all the factors that influence the outcome under study. For this design we not only have to know them but we have to be able to measure them, and we have to be able to find matched comparison subjects who share those characteristics with the exposed subjects. Therefore matched designs, although elegant in theory, are often difficult to employ in practice. Matching is discussed more fully in Chapter 6.

The designs we have described so far involve *internal control groups*, i.e. controls derived from the same source population as the exposed subjects (e.g. the same community, workplace, or medical practice). A rather weaker design uses an *external control group*, from a different source population. While the source populations for the exposed and unexposed groups are not the same, they must both relate to a common target population. Thus the health effects of asbestos could be examined by comparing workers who use asbestos with workers in the same industry who have generally similar jobs but do not use asbestos—an internal control group. The health effects of asbestos could also be assessed by comparing the death rates of workers using asbestos with the death rates for the whole population in that area or country—an external control group. If the effects are large, this may be an adequate design, but it is clearly a rather weak one. In a prospective cohort study of vegetarians, with some 17 years of follow-up, the overall mortality was much lower than that of the general population—a comparison with an external control group; in one of many internal comparisons, the mortality rate from all causes was reduced in those who consumed fresh fruit daily compared with the other members of the cohort [20]. The main options in the design of cohort studies are summarized in **Ex.** 4.11.

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025

05-Elwood-Chap04 12/20/06 5:04 PM Page 103

SOME OPTIONS IN THE DESIGN OF COHORT STUDIES

Design	Exposed group	Comparison group	Applicability
Randomized			
randomized trial	random selection from eligible population: intervention applied	random selection from same eligible population: intervention not applied	only for ethical interventions of likely benefit
Unmatched internal controls			
one or more outcomes, confounders not fully known	exposed subset of an eligible population	representative sample of unexposed members of same eligible population; no individual matching	preferable if multiple outcomes; confounders controlled in analysis
Matched internal controls			
specific outcome, main confounders known	exposed subset of an eligible population	unexposed subjects matched for other factors which influence outcome	only if outcome specified and main confounders known in advance
External controls			
	exposed subjects, or all members of a population with high exposure	all or sample of another population with no exposure or lower exposure	if internal controls not possible

Ex. 4.11. Design of cohort studies. Some methods of selection of exposed and comparison groups in cohort and intervention studies. The list is not meant to be exhaustive

to produce new results. Consider the situation faced by investigators in the mid-1960s concerning the effects of the contraceptive pill. Here was a new pharmacological preparation being used by very large numbers of women which could have major effects on their health. To show such effects, or to demonstrate their absence, required a long-term cohort study capable of assessing multiple endpoints and of giving results that would be widely applicable. Such a study would be a large, expensive, and long-term commitment, not easily repeated; therefore the design needed to optimize both internal and external validity. Two such studies were set up in the UK.

The first study (**Ex.** 4.12) was set up in 1968 by the Royal College of General Practitioners (RCGP), using patients registered with 1400 volunteer general

4.11 Design of cohort studies. Some methods of selection of exposed and comparison groups in cohort and intervention studies. The list is not meant to be exhaustive

Examples of prospective cohort studies

We will describe two studies, which, while set up some time ago, have interesting design features and may be
p. 103 regarded as classic studies. Both are continuing ↳ to produce new results. Consider the situation faced by investigators in the mid-1960s concerning the effects of the contraceptive pill. Here was a new pharmacological preparation being used by very large numbers of women which could have major effects on their health. To show such effects, or to demonstrate their absence, required a long-term cohort study capable of assessing multiple endpoints and of giving results that would be widely applicable. Such a study would be a large, expensive, and long-term commitment, not easily repeated; therefore the design needed to optimize both internal and external validity. Two such studies were set up in the UK.

The first study (**Ex.** 4.12) was set up in 1968 by the Royal College of General Practitioners (RCGP), using
p. 104 patients registered with 1400 volunteer general ↳ practitioners [21]. The general practitioners selected the first two women in each month for whom they prescribed an oral contraceptive, either for the first time or as a repeat prescription. For each, a control was selected as the next woman identified from the practice records

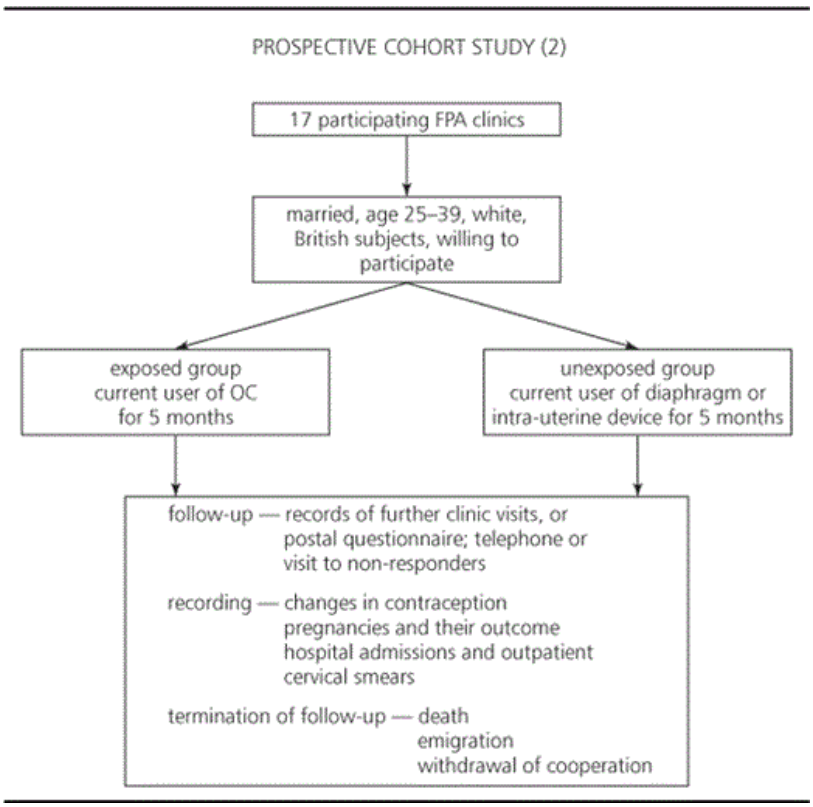
Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025

p. 105 who was aged within 3 years of the oral contraceptive user, but who had never used an oral contraceptive. ↳ Both users and non-users had to be married or living as married, and thus were likely to be sexually active. The follow-up was based on the general practitioner's regular records, including further information on oral contraceptive use, pregnancies, and related events. Patients who had left their original practitioner, or whose practitioner withdrew from the study, or who were supplied with oral contraceptives from other sources ceased follow-up at that time. This study has continued: follow-up over 25 years has shown the same overall death rate in oral contraceptive users and non-users of oral contraceptives, but increases in deaths from cervical cancer and from cerebrovascular disease, and a decrease in ovarian cancer deaths, in current and recent users [22].

This image cannot be displayed online for copyright reasons.

The other study was based on 17 of the largest clinics run by the Family Planning Association (FPA) [23] (**Ex.** 4.13). Eligible subjects had to be married, aged 25–39, a white British subject, and be willing to participate;
p. 106 these criteria were primarily to ensure adequate follow-up. Oral contraceptive users were ↳ defined by current and past use over 5 months, and the unexposed group was defined as women using a diaphragm or an intra-uterine device for at least 5 months without prior exposure to oral contraceptives. The 5-month duration criterion was to eliminate substantial numbers of women who would change their method of contraception only a few months after starting. Follow-up information was based on the FPA clinic records, but if no further appointments were recorded, a follow-up form was sent directly to the patient, supported by telephone calls or home visits where necessary. Patients were asked on recruitment to give the name of their family doctor and of two contact persons to assist in follow-up. Information about hospital visits was sought on both clinic and direct mail follow-up forms; the primary outcome measures used in analysis were mortality and morbidity recorded as inpatient or outpatient visits. Follow-up ceased at death, emigration from the UK, or at the subject's request. This study has also continued: follow-up to the year 2000 showed no effect of oral contraceptive use on total mortality, increased deaths from cervical cancer, and, in agreement with the RCGP study, a decreased death rate from ovarian cancer [24].

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025



4.13 Design of a prospective cohort study of oral contraceptive (OC) use: based on Family Planning Clinics. From Vessey *et al*. [23]

In both these studies, the prime considerations were to achieve high follow-up and high internal validity, after choosing source populations which gave reasonable external validity. In the FPA study, the loss to follow-up was only around 0.7 per cent per year, and was similar in the different contraceptive groups. This good follow-up was achieved by the eligibility criteria which tended to select women with a stable lifestyle, and applied to both the oral contraceptive and comparison groups. This advantage in internal validity was achieved at the cost of some external validity. The women in the FPA study are not representative of all oral contraceptive users in the UK, and exclude, for example, women of non-white origin and younger unmarried women who may have different sexual behaviours. However, any major biological associations assessed in this study might well apply to other groups. A greater limitation is that the comparison was between oral contraceptive users and women using a different method of contraception, so that any differences in a particular outcome may be due to either the oral contraceptive or the other method. However, it was useful that the comparison group comprised two major subgroups, users of a diaphragm or of an intrauterine device. Thus in one study it was shown that the frequency of cervical cancer and dysplasia was lower in women who used a diaphragm than in either of the other two groups, suggesting a protective effect of diaphragm use rather than an increased risk from the other methods. To assess whether cervical cancer was increased in oral contraceptive users, the relevant comparison was between oral contraceptive users and users of an intra-uterine device, excluding diaphragm users entirely from that analysis [25].

p. 107 The external validity of the RCGP study may be somewhat greater, as the criteria are looser, and the users and non-users chosen may be representative of all users in a particular practice. However, the number of eligible women who were selected but did not choose to enter the study was not recorded. The non-users included women who were not using any method of contraception, which is probably not particularly beneficial, as differences between them and oral contraceptive users in regard to other features related to sexual activity may be substantial. However, the participating general practitioners are unlikely to be representative of all

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025

practitioners, and therefore in this study the women cannot be regarded as a representative sample of British oral contraceptive users.

Both these designs fail to fulfil one of the criteria set out in Ex. 4.5; the exposed group were not defined from the time of first exposure, but were identified as a prevalent sample of oral contraceptive users. The recruitment process would have been much more difficult if women had to be recruited at first use only, as many women would stop oral contraceptive use after only a short time and therefore contribute little to the study. As a result, neither of these studies is powerful in the assessment of short-term effects of oral contraceptive use, as women who started oral contraceptives and had ill effects immediately would be under-represented in both studies. In both studies, a degree of external validity has been sacrificed to facilitate follow-up and achieve good internal validity. The ways in which the data were collected in these studies will be reviewed in Chapter 5.

Retrospective cohort studies

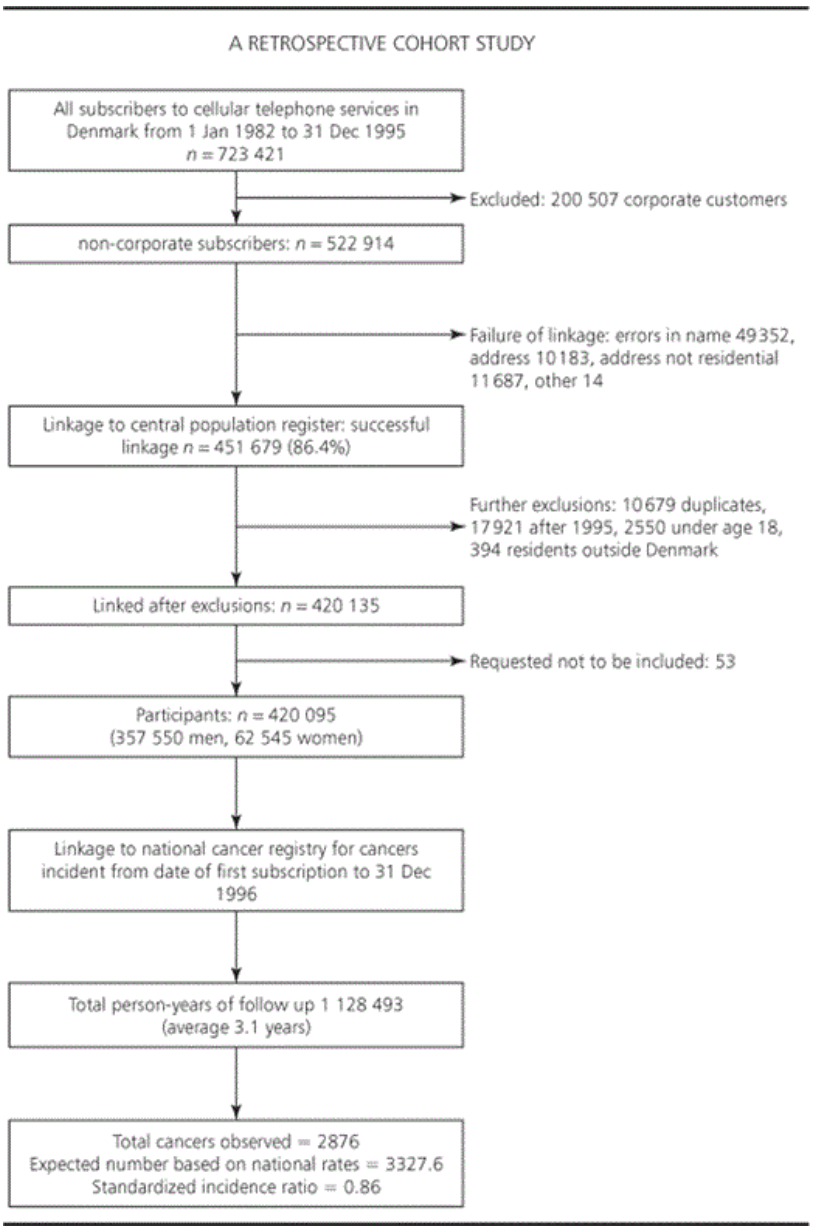
The prospective cohort studies described above are clearly very major undertakings, requiring many years of follow-up to produce results. If the essential information for a particular study can be obtained from records that already exist, the advantages of a cohort study can be exploited without the need for the long length of time required for prospective follow-up. Many such studies have been done on groups of people who can be identified as sharing an important exposure in the past, such as occupational groups. Very powerful studies can be carried out efficiently using computer-based record linkage techniques if high-quality databases are available.

Example of a retrospective cohort study

A good example of a retrospective cohort study is a study set up to assess possible links between the use of cellular (mobile) telephones and cancer in Denmark [26] (**Ex.** 4.14). Cell phone services in Denmark started in
p. 108 1982 and were provided by two operating companies. After scientific and ethical review, ↳

p. 109 ↳
both companies made their records available for the study, and it was possible to link the lists of subscribers to a central population register which carried information on mortality and date of emigration, and in turn to link this to the Danish cancer registry which records information on all cancers diagnosed. This demonstrates the immense potential for research of the use of linkable population registries, which here include a unique 10-digit personal identification number that assists accurate linkage between the registers. In many other countries such links are impossible either because the registries do not exist, or because legal issues prevent the linkages being established; in fact, a similar proposed study using subscriber lists for cell phone customers in the USA had to be abandoned after a legal challenge.

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025



4.14 Design of a retrospective cohort study: a national study of cancer occurrence in mobile phone subscribers in Denmark. From Johansen *et al*. [26]

Thus the investigators were able to identify all 723 421 subscribers to cell phone accounts, from which they excluded 200 507 corporate customers, leaving 522 914 non-corporate subscribers. A key assumption is that the non-corporate subscriber is the user of the cell phone, which is obviously not always true. Further exclusions were because of errors in name or address of the telephone user, addresses which were not residential, duplicates, subscriptions initially identified but which were outside the eligibility period chosen, persons under the age of 18, and residents in Greenland and the Faroe Islands. The study needed only access to the registers, and the subjects were not asked for consent and were not informed if they were included. The study was publicly announced, and the two telephone companies issued a notice that subscribers could contact them if they wished to be excluded; only 53 people actively excluded themselves in this way. The final cohort was of 420 095 cellular telephone subscribers, representing 80.3 per cent of the non-corporate subscribers initially identified. Linkage to the cancer registry then allowed the calculation of incidence rates for various cancers, which were compared with the incidence in the whole country as the comparison population. Using information from the cell phone service providers, the data could be analysed by the year of first subscription, age at first subscription, the type of telephone system, and the duration of subscription. Special attention was

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025

paid to tumours of the brain, which were analysed in detail with respect to site within the brain and pathological subtype. The results showed no indication of an increased risk of cancer in total or in any tumour type, even in subjects with the longest duration of exposure, which was the only measure of intensity of exposure in the study.

The design makes this study strong in that the participants are likely to be representative of all subjects in Denmark using cellular phones, although it has been pointed out that the exclusion of corporate subscribers
p. 110 could exclude subjects with the most intense usage, such as sales people. The weaknesses of ↳ the study are that there is no information on the individual extent of use or, for example, on which side of the head the cell phone was normally used, which has been assessed in case–control studies using questionnaires, and the short latency period. In this study the eligible cell phone subscriptions started between 1982 and 1995, and cancer incidence was analysed for each subscriber from the start of their subscription up to 1996. Thus the study includes those in whom the occurrence of cancer would have been shortly after their first subscription, and is limited to a maximum of 14 years between first cell phone use and cancer incidence. The mean follow-up time was only 3 years. This is too short to demonstrate a classic cancer incidence effect, although it could detect a promotion effect, and this limitation is acknowledged by the authors. A strength of the study is its precision, as the study is extremely large. For all cancers, there were 2876 cancers in men compared with 3327.6 expected, giving a relative risk of 0.86, with statistical 95 per cent confidence limits of 0.83–0.90. These are explained later in Chapter 7, but illustrate the precision of this result. On the other hand, for the tumours of greatest interest, those of the brain and nervous system, the numbers were much more modest, there being 135 observed cases compared with 142.8 expected giving a risk ratio of 0.95 with considerably wider confidence limits of 0.79–1.12; the study does not confidently exclude a 12 per cent increase in risk for these tumours.

Selection of subjects for a case–control study

We will now look at the selection of subjects for a case–control study, which follows very similar principles to those outlined in Ex. 4.5. Here we are selecting groups on the basis of outcome—selecting a case group who have already suffered the outcome and a control group.

The cases should truly be cases. The inclusion of some individuals who do not in fact have the outcome in question within the case group will tend to dilute the case group and bias the results of the study towards the null value. However, as we have seen earlier, this ideal has to be balanced with the logistical difficulties of ascertaining a representative case group. Suppose that the diagnosis of the disease requires complex procedures. Then, to ensure that all those classified as cases do in fact have the disease may involve restricting the study to subjects who have had the opportunity to go through such diagnostic tests. This will exclude individuals who have the disease but have not been investigated so thoroughly. The restricted case series may not be representative of the disease in the wider community; it may be slanted towards individuals with more severe or more manifest disease. The dilution effect of including some non-cases within the case series has to
p. 111 be balanced against the possible non-representativeness of a limited case series. It may be helpful to categorize ↳ cases in terms of the certainty of their definition; thus in a case–control study of venous embolism and hormone replacement therapy, the association seen was stronger for the cases with a definite rather than a possible diagnosis [27].

Ideally, the cases should be recently diagnosed. A series of prevalent cases, such as all cases currently being seen in a clinic or existing in a community, will exclude those subjects who have developed the disease and then left the area, died, or recovered. Such subjects will be different in a number of ways from those who still have the disease and therefore are included in the sample. In studies of the outcome of disease in groups of subjects seen in hospital, a frequent error is to study only those subjects who are still under follow-up by the hospital, rather than all patients diagnosed with the disease, irrespective of whether they are being followed up

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025

or not. The patients not under follow-up include those with particularly bad outcomes, who may have died or been admitted elsewhere, and sometimes those with particularly good outcomes, who need not return for further care.

A major issue in case–control studies of disease is the choice between a case series chosen from one or more hospitals, or one derived from a community. Hospital series are acceptable if a very high proportion of those developing the disease will come into hospital for diagnosis or treatment. If that is not so, the hospital-based cases may differ substantially from those in the community. This restriction may be accepted in view of the logistic advantages of basing a study on hospital cases, but considerable care is then needed in the generalization of the results.

Choice of the control group in case–control studies

Choosing control groups in case–control studies is more complicated than in cohort studies. In the most frequently used design, subjects in the control group are chosen to be truly without the outcome of interest at the time the study is performed. The objective of this design is that by comparing representative samples of cases and of controls who do not have the outcome, the study will provide an estimate of the odds ratio in the underlying population. Misclassification, i.e. including some who are actually cases in the control group, will have a dilution effect and bias the results of the study towards the null value. If the prevalence of the outcome under study is small (such as cancer), only a few controls are likely to have the disease and the effect of this misclassification will usually be too small to be important. Where the outcome is relatively common in the underlying source population (such as hypertension or depression), this effect could be considerable, and it may be necessary to assess potential controls to exclude disease. The benefits of doing this must be compared with the likely fall in participation rate produced by such an assessment, added to the logistic, cost, and ethical issues.

p. 112

However, as noted in Chapter 3, there is an alternative design for a case–control study, in which the controls are selected not to be representative of subjects without the case condition, but designed to be representative of the eligible population at risk. In this design, the format of the results directly produces a relative risk estimate, and a case subject is also eligible for sampling as a control. On this basis, because relative risk is being estimated, there is no issue of misclassification in terms of the controls.

Exhibit 4.15 illustrates some of the designs that are often used for case–control studies. A primary choice is between selecting cases and controls from the community, or from a health care facility or similar ‘institutional’ source. One strong design uses a case series that is a total or representative sample of all affected subjects drawn from a specified source population, and a control group that is chosen as a representative sample of unaffected members of that same source population. The source population may be a community or a health care source.

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025

SOME OPTIONS IN CASE–CONTROL STUDIES

Design	Case group	Control group	Applicability
Unmatched community-based	all or representative sample of all affected subjects in source population	representative sample of unaffected (or of all) members of same source population; no individual matching	preferable for multiple exposures or if confounders not known; confounders controlled in analysis
Unmatched institution-based	all or representative sample of all affected subjects in eligible population	sample of unaffected members of same eligible population; no individual matching	preferable for multiple exposures or if confounders not known; confounders controlled in analysis
Matched community- or institution-based	affected members of eligible population	unaffected subjects chosen to be similar to cases on certain specified matching factors; from same eligible population	only if exposure specified and main confounders known in advance

4.15 Design of case–control studies. Some methods of selection of case and control groups in case–control studies. The list is not meant to be exhaustive

p. 113 A community-based design has the advantage that if all cases of the disease of interest can be ascertained in the community, the case series can be fully representative. Moreover, the information from the control group may be much easier to interpret, as the controls will be healthy subjects representative of that community. The source population is also more closely related to other target populations, which will make the further generalization of the results more straightforward.

In a study based on a health care facility, for example comparing patients with a particular disease with patients with other conditions in the same hospital, a danger is that these other conditions may be related to the exposure under consideration. A useful protection with hospital-based case–control designs is to ensure that the control subjects are selected with a range of other diagnoses, as it is unlikely that the exposure under consideration will be related to all of them. Patients with diagnoses likely to be associated (positively or negatively) with the exposure factor under assessment should not be eligible as controls. Thus in a case–control study of venous embolism and hormone replacement therapy, data were presented for nine diagnostic categories of hospital controls, showing considerable variation in the frequency of use of hormone replacement therapy [27]. Also, the applicability of the results to the target population may be more difficult to assess in a hospital-based study.

From the above, the advantages of community-based case–control studies would seem to be considerable, but against these must be balanced the greater difficulty of carrying out such studies, and particularly of ensuring a high response rate in the control series. It is more difficult to obtain a high degree of cooperation from subjects in the community as they have less incentive to be involved in the study than have patients who have been treated. Further, some studies may require clinical information on comparison subjects that may not be easy to obtain from subjects chosen from the community.

It is usually valuable to achieve general comparability between cases and controls by balancing the numbers chosen in terms of gender, age group, and perhaps a few other demographic factors, such as place of residence

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025

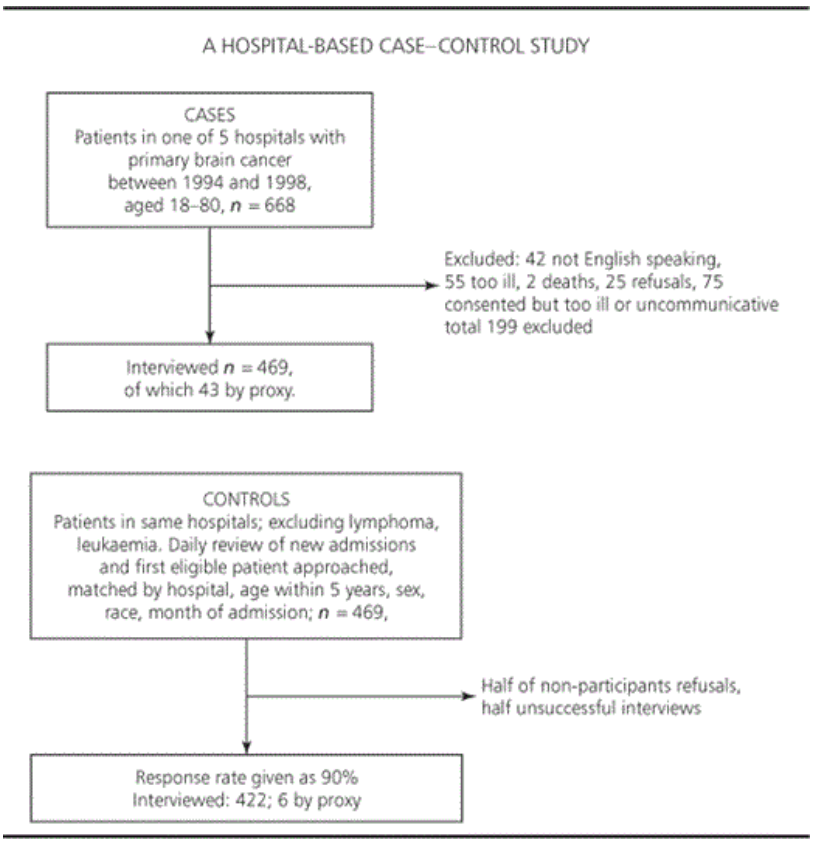
or hospital. This is referred to as *frequency matching* and is mainly for efficiency, as will be discussed in Chapter 6. Controls can instead be chosen to be *individually matched* to the case series with regard to specified confounding factors, which are associated with both outcome and exposure. This is useful under certain circumstances, discussed in Chapter 6, but adds complexity to the study. In an individually matched design, the matching of the control subjects takes precedence over their other characteristics, but beyond this, institutional or community sources of controls may be used.

p. 114 Finally, the control subjects need to be chosen so that the information on exposure can be obtained in a similar manner as in the case group. This may involve modification of the case or the control criteria.

Examples of case–control studies

The association between cell phone use and brain cancer has also been assessed in several case–control studies, such as that by Muscat *et al*. [28] (**Ex.** 4.16). This uses a standard hospital-based case–control design to enrol cases and controls efficiently for what is a fairly rare disease. The cases were defined as patients aged 18–80 years who had been diagnosed with a primary brain cancer during the previous year, and were attending
p. 115 one of five hospitals in New York, Providence, and Boston. Controls were chosen as inpatients in the ↳ same hospital, selected by checking each day's admissions and selecting the first patient who met the eligibility criterion of English language use and were of a suitable age, sex, race, and month of admission. These characteristics were set so that the control groups would eventually be similar in distribution on these factors to the case group, although they were not individually matched. Because there had been previous reports of links between radiofrequency exposures and lymphoma and leukaemia, patients with these conditions were excluded from the control group. The controls had diagnoses that fell into five major categories, and the use of cell phones is described for these five categories in the study results. This does suggest some variability, such as a lower rate of cell phone use in controls who had other cancers. Much of this difference is probably due to age, but it does suggest that selecting controls with a narrow range of diagnoses could be misleading. Cases and controls were then interviewed while in hospital by health professionals. This system had the advantage of giving high response rates. Of the 668 patients with brain tumours identified, 42 were not English-speaking, two had died, 55 were thought to be too ill to be approached, 75 agreed but were too ill to participate, and 25 declined to participate, giving a participation rate of 469/668 (70 per cent). The authors describe 571 subjects as eligible (omitting the 42 non-English-speakers and the 55 who were too ill), giving the response rate as 82 per cent (469/571). In the control subjects the response rate is given as 90 per cent. The high response rate is an advantage in this study. However, there are questions about whether the control group is representative of the general population in terms of cell phone exposure, and the validity of the information could be compromised as the interviewers were aware of the diagnosis of the subjects.

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025



4.16 Selection of cases and controls in a hospital-based case–control study assessing the association between the use of cellular phones and the risk of brain cancer. From Muscat *et al*. [28]

In contrast, in a case–control study of breast cancer in New Zealand assessing primarily oral contraceptive use, the investigators concentrated on obtaining a population-based series of both cases and controls [29] (**Exs** 4.17 and 4.18). This was a complex process. The cases were defined as all New Zealand resident women aged 25–54 who were diagnosed with breast cancer between 1983 and 1985; they could be identified from cancer registries. To obtain a representative population sample as a control series was more difficult. The most complete available population record was the electoral register, but this does not cover the entire population and does not give age. It was decided to collect the information by a telephone interview using a small number of trained interviewers at one site. Therefore both cases and controls had to have a listed telephone number, and the cases had to be on the electoral register to be comparable to the controls. Of 739 women identified within the age
p. 116 range from the cancer registers, 189 were excluded by not being on the electoral roll, not ↳

p. 117 ↳
having a listed telephone number, or having had previous breast cancer, giving 550 eligible subjects. The approach to the breast cancer patients required consent from their doctor, fitness to be interviewed, an interview not more than 8 months after diagnosis, and the patient's consent. There were 433 women who participated, representing 79 per cent of the 550 eligible subjects, but only 59 per cent of the original source population. It is fairly typical of modern epidemiology that the least of the investigators' problems was failure of the subjects to agree to interview, with only 14 refusals; more women were excluded because their doctors did not give permission for them to be approached.

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025

THE CASE SERIES IN A CASE–CONTROL STUDY	
Source population	all New Zealand women aged 25–54 diagnosed with breast cancer between 1 July 1983 and 30 June 1985
Eligible population	women aged 25–54, with histologically confirmed breast cancer notified to the NZ National Cancer Registry or to the Auckland Breast Cancer Study group between above dates ($n = 739$); no previous breast cancer; on current electoral roll; whose telephone number was found;
$n = 550$	exclusions = 189, eligible population = 550
Participant population	had to have permission from their physician (28 not given) be identified in time to allow interview 4–8 months from diagnosis (49 too late) still alive (8 had died) well enough (4 too ill) exclusions (14 other exclusions) agreed to participate (14 refused)
$n = 433$	participant population 433; exclusions 117 all participants gave usable information on oral contraceptive use
Participants/eligible = 78.7% Participants/source = 58.6%	

4.17 A case–control study. The source, eligible, and participant populations in the case series of a case–control study of breast cancer in New Zealand. From Paul *et al*. [29]

THE CONTROL SERIES IN A CASE–CONTROL STUDY	
Source population	all New Zealand women aged 25–54 and without diagnosed breast cancer
Eligible population	women on electoral register, aged 25–54, with a telephone number, with no history of breast cancer: selected by
$n = 1110$	random sampling from register; estimate of number = 1110
Participant population	still alive (10 had died) well enough (4 too ill) in New Zealand (12 overseas) no language difficulty (13 excluded) agreed to participate (99 refused) traced (75 not traced, of whom some were likely to be outside the age range)
$n = 897$	participant population 897; exclusions 213
Participants/eligible = 80.8% Participants/source = unknown	

4.18 A case–control study. The source, eligible, and participant populations in the control group in a case–control study of breast cancer in New Zealand. From Paul *et al*. [29]

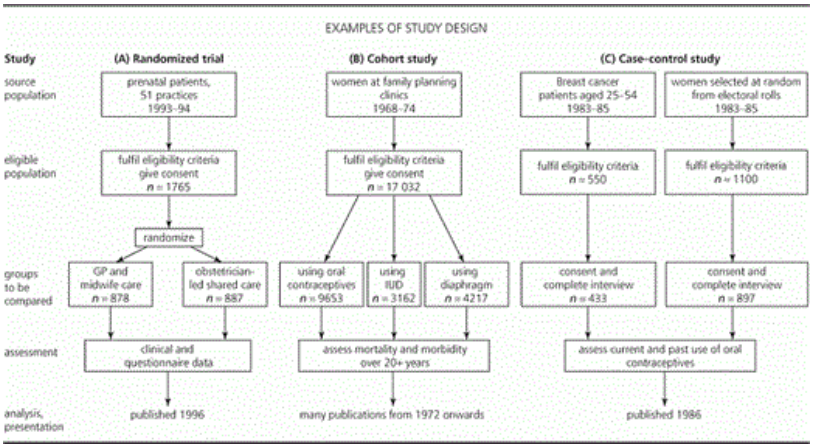
The control group sampling was further complicated by the lack of age information. The investigators had to take a random sample of women from the electoral rolls, exclude those for whom a telephone number could not be found, and then write to the women asking for their participation; they could determine age only if she responded, and previous breast cancer only when the interview was carried out. The eligible population cannot be precisely determined, but the best estimate is 1110 and the estimated participation rate is 81 per cent. This is a minimum estimate, as amongst the women not traced there are likely to be some who would have been ineligible because of age. Exclusions of controls because of illness or death were fewer than for cases, and of course a doctor's permission was not required. The voluntary response rate of eligible controls was 897/996 (90 per cent); this is lower than that of the cases, although is still very high. The lower response would be expected as the control subjects have less motivation to take part in a health study than the case subjects who have had a serious disease.

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025

In both these designs, the emphasis in detailed design is, as it should be, on setting eligibility criteria and exclusion processes that are the same for the case and the control series. Both studies have considerable difficulties that could compromise the results, and it is not clear whether a hospital design is inherently better or worse than a community design. It is likely for instance that the case series interviewed in the brain cancer study represents a more complete series of cases than the case series interviewed in the breast cancer study. However, the control series in the breast cancer study is more likely to be representative of women in the general community.

Comparison of study designs

Three of the studies reviewed here are summarized in **Ex.** 4.19, which shows for each where the group of prime interest and the comparison group diverge in terms of the participant, eligible, and source populations. The potential for substantial differences between the groups being compared clearly increases as we go from the randomized trial design to the case–control study.



4.19 Example of study design: (A) randomized clinical trial (Tucker et al. [18]); (B) prospective cohort study (Vessey *et al*. [23]); (C) case–control study (Paul *et al*. [29])

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025

Self-test questions (answers on p. 495)

Q4.1 Define the target, source, eligible, and participant populations in the following study. To assess the role of magnetic fields in causing childhood leukaemia, children with leukaemia treated in a major referral centre were identified; those in a terminal stage of illness were excluded, and others were interviewed with a 60 per cent response.

Q4.2 Suppose in the study just described, an association is found with a history of measles in the first year of life (odds ratio, 2.5). Summarize the concepts of internal and external validity with regard to this result.

Q4.3 What effects can selection bias have on the results of a study?

Q4.4 In selecting a case series for a case–control study, 1000 subjects with the disease in question are identified, and 900 fulfil the eligibility criteria of disease categorization and age. Of these, address information is incomplete on 60, and the doctors of 100 do not give permission for them to be approached. All the remaining subjects are approached for interview, and 500 consent; however, 10 per cent of those interviewed have missing data on the key variables for the analysis. Summarize the selection process, calculating the participation rate, voluntary response rate, and ratio of the participants to the eligible and the source populations.

Q4.5 In a randomized trial of smoking cessation, smokers are randomly allocated to be offered an intervention or not. After 1 year, the frequency of smoking cessation in those randomized to no intervention was 20 per cent. Of those randomized to the intervention group, only half accepted the intervention, and their cessation rate was 50 per cent. The cessation rate in those randomized to the intervention but who did not accept the programme offered was 10 per cent. What is the most appropriate summary result from this study?

Q4.6 What four criteria should be fulfilled by the exposed group in a cohort study?

Q4.7 In a case–control study, cases are identified through general practitioners (family doctors) and interviewed by telephone. What selection principles apply to the controls?

Q4.8 What is meant by a single-blind or double-blind trial?

Q4.9 In the context of a cohort study of workers exposed to a particular chemical, how could an internal and an external control group be defined?

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025

References

1. Slanetz PJ, Moore RH, Hulka CA, *et al*. Physicians' opinions on the delivery of mammographic screening services: immediate interpretation versus double reading. *AJR Am J Roentgenol* 1996; **167**(2): 377–379. 10.2214/ajr.167.2.8686608
WorldCat Crossref

2. Ellenberg JH, Nelson KB. Sample selection and the natural history of disease. Studies of febrile seizures. *JAMA* 1980; **243**: 1337. 10.1001/jama.243.13.1337
WorldCat Crossref

3. The Toronto Leukemia Study Group. Results of chemotherapy for unselected patients with acute myeloblastic leukaemia: effect of exclusions on interpretation of results. *Lancet* 1986; **i**: 786–788.
WorldCat

4. Morris JN. *Uses of Epidemiology* (1st edn). Edinburgh: Livingstone, 1957.
Google Scholar Google Preview WorldCat COPAC

5. Ziel HK, Finkle WD. Increased risk of endometrial carcinoma among users of conjugated estrogens. *N Engl J Med* 1975; **293**: 1167–1170. 10.1056/NEJM197512042932303
WorldCat Crossref

6. Hulka BS, Grimson RC, Greenberg BG, *et al*. 'Alternative' controls in a case–control study of endometrial cancer and exogenous estrogen. *Am J Epidemiol* 1980; **112**(3): 376–387.
WorldCat

7. Morton LM, Cahill J, Hartge P. Reporting participation in epidemiologic studies: a survey of practice. *Am J Epidemiol* 2006; **163**(3): 197–203. 10.1093/aje/kwj036
WorldCat Crossref

8. Edwards P, Roberts I, Clarke M, *et al*. Increasing response rates to postal questionnaires: systematic review. *BMJ* 2002; **324**: 1183. 10.1136/bmj.324.7347.1183
WorldCat Crossref

9. Shapiro S, Venet W, Strax P, Venet L, Roeser R. Ten- to fourteen-year effect of screening on breast cancer mortality. *J Natl Cancer Inst* 1982; **69**: 349–355.
WorldCat

10. Multiple Risk Factor Intervention Trial Research Group. Multiple risk factor intervention trial: risk factor changes and mortality results. *JAMA* 1982; **248**: 1465–1477. 10.1001/jama.248.12.1465
WorldCat Crossref

11. Fletcher SW, Black W, Harris R, Rimer BK, Shapiro S. Report of the International Workshop on Screening for Breast Cancer. *J Natl Cancer Inst* 1993; **85**: 1644–1656. 10.1093/jnci/85.20.1644
WorldCat Crossref

12. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. *JAMA* 2003; **290**(12): 1624–1632. 10.1001/jama.290.12.1624
WorldCat Crossref

13. Holmberg L, Anderson H. HABITS (Hormonal Replacement Therapy After Breast Cancer–Is It Safe?), a randomised comparison: trial stopped. *Lancet* 2004; **363**: 453–455. 10.1016/S0140-6736(04)15493-7
WorldCat Crossref

14. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025

of parallel-group randomised trials. *Lancet* 2001; **357**: 1191–1194. 10.1016/S0140-6736(00)04337-3
WorldCat Crossref

15. Altman DG, Schulz KF, Moher D, *et al*. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001; **134**: 663–694.
WorldCat

16. Moher D, Jones A, Lepage L. Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *JAMA* 2001; **285**: 1992–1995. 10.1001/jama.285.15.1992
WorldCat Crossref

17. Seaton ED, Charakida A, Mouser PE, Grace I, Clement RM, Chu AC. Pulsed-dye laser treatment for inflammatory acne vulgaris: randomised controlled trial. *Lancet* 2003; **362**: 1347–1352. 10.1016/S0140-6736(03)14629-6
WorldCat Crossref

18. Tucker JS, Hall MH, Howie PW, *et al*. Should obstetricians see women with normal pregnancies? A multicentre randomised controlled trial of routine antenatal care by general practitioners and midwives compared with shared care led by obstetricians. *BMJ* 1996; **312**: 554–559. 10.1136/bmj.312.7030.554
WorldCat Crossref

p. 121

19. Baird PA, Yee IML, Sadovnick AD. Population-based study of long-term outcomes after amniocentesis. *Lancet* 1994; **344**: 1134–1136. 10.1016/S0140-6736(94)90635-1
WorldCat Crossref

20. Key TJA, Thorogood M, Appleby PN, Burr ML. Dietary habits and mortality in 11 000 vegetarians and health conscious people: results of a 17 year follow up. *BMJ* 1996; **313**: 775–779. 10.1136/bmj.313.7060.775
WorldCat Crossref

21. Royal College of General Practitioners. *Oral contraceptives and Health: An Interim Report from the Oral Contraception Study of the Royal College of General Practitioners*. New York: Pitman Medical; 1974.
Google Scholar Google Preview WorldCat COPAC

22. Beral V, Hermon C, Kay C, Hannaford P, Darby S, Reeves G. Mortality associated with oral contraceptive use: 25 year follow up of cohort of 46 000 women from Royal College of General Practitioners' oral contraception study. *BMJ* 1999; **318**: 96–100. 10.1136/bmj.318.7176.96
WorldCat Crossref

23. Vessey M, Doll R, Peto R, Johnson B, Wiggins P. A long-term follow-up study of women using different methods of contraception-an interim report. *J Biomed Sci* 1976; **8**: 373–427.
WorldCat

24. Vessey M, Painter R, Yeates D. Mortality in relation to oral contraceptive use and cigarette smoking. *Lancet* 2003; **362**: 185–191. 10.1016/S0140-6736(03)13907-4
WorldCat Crossref

25. Vessey MP, Lawless M, McPherson K, Yeates D. Neoplasia of the cervix uteri and contraception: a possible adverse effect of the pill. *Lancet* 1983; **ii**: 930–934. 10.1016/S0140-6736(83)90451-8
WorldCat Crossref

26. Johansen C, Boice JD, Jr., McLaughlin JK, Olsen JH. Cellular telephones and cancer—a nationwide cohort study in Denmark. *J Natl Cancer Inst* 2001; **93**: 203–207. 10.1093/jnci/93.3.203
WorldCat Crossref

27. Daly E, Vessey MP, Hawkins MM, Carson JL, Gough P, Marsh S. Risk of venous thromboembolism in users of hormone replacement therapy. *Lancet* 1996; **348**: 977–980. 10.1016/S0140-6736(96)07113-9

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025

WorldCat Crossref

28. Muscat JE, Malkin MG, Thompson S, *et al*. Handheld cellular telephone use and risk of brain cancer. *JAMA* 2000; **284**: 3001–3007. 10.1001/jama.284.23.3001
WorldCat Crossref

29. Paul C, Skegg DCG, Spears GFS, Kaldor JM. Oral contraceptives and breast cancer: a national study. *BMJ* 1986; **293**: 723–
p. 122 726. ↳ 10.1136/bmj.293.6549.723
WorldCat Crossref

Downloaded from https://academic.oup.com/book/16544/chapter/172342337 by INACTIVE user on 22 September 2025