CHAPTER 1

Fundamental concepts

This chapter provides a summary background to observational studies, their main purposes, the common types of designs, and some key design features. Further details on design and analysis are illustrated using examples in later chapters, and from other textbooks [1–3].

1.1 Observational studies: purpose

Two distinct study designs are used in medical research: **observational** and **experimental**. Experimental studies, commonly called clinical trials, are specifically designed to intervene in some aspect of how the study participants live their life or how they are treated in order to evaluate a health-related outcome. A key feature of a clinical trial is that some or all participants receive an intervention that they would not normally be given. Observational studies, as the term implies, are not intentionally meant to intervene in the way individuals live or behave or how they are treated.[#] Participants are free to choose their lifestyle habits and, with their physician, decide which interventions they receive when considering preventing or treating a disorder. Box 1.1 shows the most common purposes of observational studies.

1.2 Specifying a clear research question: exposures and outcomes

The **research question(s)**, which can also be referred to as **objectives**, **purpose**, **aims**, or **hypotheses**, should be clear, easy to read, and written in non-technical language where possible. They are usually developed to address a research issue that has not been examined before, to corroborate or refute previous evidence, or to examine a topic on which prior evidence has had shortcomings or been scientifically flawed.

[#]Though in reality, just by being in a study could alter someone's behaviour or lifestyle habits.

A Concise Guide to Observational Studies in Healthcare, First Edition. Allan Hackshaw.

Box 1.1 Common purposes of observational studies

- Examine the opinions of a single group of people on a health-related topic(s)
- Describe the health-related characteristics (e.g. demographics, lifestyle habits, genes, biological measurement, or imaging marker) of a single group of people
- Estimate the occurrence of a disorder at a given time, or trends over time
- Examine features of a disorder (e.g. how it affects patient's lives, how they are managed/treated, and short- or long-term consequences)
- Find associations between the health-related characteristics among a single group of people or across two or more groups
- Examine risk factors (including casual factors) for a disorder or early death
- Examine prognostic factors (i.e. those that can predict the occurrence of a disorder or death from the disorder)
- Evaluate a healthcare intervention for prevention or treatment

Find new scientific information
Plan the use of future resources
Change public health education, policy, or practice
Change clinical practice

Disease prevention, detection, or treatment

There is a distinction between **objectives** and **outcome measures** (or **endpoints**). An outcome measure is the specific quantitative measure used to address the objective. For example, a study objective could be 'to examine the smoking habits of adults'. Possible corresponding endpoints could be either 'the proportion of all participants who report themselves as smokers' or 'the number of cigarettes smoked per day', but they are quite different endpoints. Box 1.2 shows examples of objectives and outcome measures.

It can be easy to specify the research question or objective for studies that involve simply describing the characteristics of a single group of people (e.g. demographics, or biological or physical measurements). For example:

- What proportion of pregnant women give birth at home?
- What is the distribution of blood pressure and serum cholesterol measurements among men and women aged over 50?

Box 1.2 Examples of objectives and outcome measures (endpoints)

Objective	Outcome measure
To examine the effectiveness of statin therapy in people with no history of heart disease	Mean serum cholesterol level
To evaluate blood pressure as a risk factor for stroke	The occurrence (incidence) of stroke
To examine the smoking and alcohol drinking habits of medical students	The number of cigarettes smoked per day and the number of alcohol units consumed in a week
To determine whether there is an association between arthritis and coffee consumption	The occurrence of arthritis
To examine the association between age and blood pressure	Age and blood pressure measured on every subject

- Are patients satisfied with the quality of care received in a cancer clinic?

Clinical trials often have a single primary objective, occasionally two or three at most, each associated with an endpoint. However, there can be more flexibility on this for observational studies unless they have been designed to change a specific aspect of public health policy. Many observational studies have several objectives, some of which may only arise during the study or at the end, and they can also be exploratory.

Examining the effect of an exposure on an outcome

While some researchers seek only to describe the characteristics of a single group of people (the simplest study type), it is common to look at associations between two factors. Many research studies, both observational studies and clinical trials, are designed to:

Examine the effect of an *exposure* on an *outcome*

Box 1.3 gives examples of these. To evaluate risk factors or causes of disease or early death, an outcome measure must be compared between two groups of people:

1. Exposed group
2. Unexposed group

Box 1.3 Examples of studies examining the effect of an exposure on an outcome

	Exposure	Outcome*
Exposures (characteristics) that cannot be changed or modified	Age	Heart disease
	BRCA1/BRCA2 gene	Breast cancer
	Family history	Alzheimer's disease
	Prostate specific antigen	Prostate cancer
	Burn size after an accident	Mortality
Exposures (characteristics) that can be changed or modified	Alcohol	Arthritis (gout)
	Frequent mobile phone use	Brain cancer
	Working with asbestos	Mesothelioma
	Body weight	Diabetes
Interventions	A new diet for obese people	**Body weight**
	Epileptic drugs during pregnancy	Birth defect
	Being treated in A&E at weekends	Death within 7 days

A&E, accident and emergency department.
Body weight is highlighted to show that a factor can be either an outcome or exposure, depending on the research question:
'What is the effect of body weight on the risk of developing diabetes?'
'What is the effect of a new diet on body weight'

*The risk of developing the specified disorder, except body weight which is a continuous measurement so there is no direct concept of risk.

An exposure is often thought to be a factor that can be avoided or removed from our lives, such as a lifestyle habit or something encountered at work or in the environment, but it can be any of the following:

- Physical or clinical characteristic
- Gene or genetic mutation
- Biomarker (measured in blood, saliva, or tissue)
- Imaging marker
- Intervention for prevention or treatment

Also, a factor can be either an exposure or an outcome, depending on the research question (e.g. body weight in Box 1.3). Considering a research study in the context of examining the relationship between exposures and outcomes greatly helps to understand the design and analysis.

"Make everything else the same": natural variation, confounding, and bias

An important consideration for all observational research studies is **variability** (**natural variation**). For example, smoking is a cause of lung cancer, but why do some people who have smoked 40 cigarettes a day for most of their adult lives not develop lung cancer, while others who have never smoked do? The answer is that people **vary**. They have different body characteristics (e.g. weight and blood pressure), different genetic make-up, and different lifestyles (e.g. diet, and exercise). People react to the same exposure in different ways.

When an association (risk or causal factor)[#] is evaluated, it is essential to consider if the observed responses are consistent with natural variation or whether there really is an effect. Allowance must be made for variability in order to judge how much of the association seen at the end of a study is due to natural variation (i.e. chance) and how much is due to the effect of the risk factor of interest. The more variability there is, the harder it is to detect an association. Highly controlled studies (such as laboratory experiments or randomised clinical trials) have relatively less variation because the researchers have control over how the study subjects (biological samples, animals, or human participants) are selected, managed, and assessed.

The best way to evaluate the effect of an exposure on an outcome is to 'make everything the same', in relation to the characteristics of the two (or more) groups being compared except the factor of interest. For example, to examine whether smoking is a cause of lung cancer, the risk of lung cancer between never-smokers and current smokers must be compared; to evaluate statin therapy for treating people with ischaemic heart disease, survival times between patients who did and did not receive statins are compared. Ideally, the exposed and unexposed groups should be identical in terms of demographics, physical and biological characteristics, and lifestyle habits, so that the only difference between the groups is that one is exposed to the factor of interest (smokes or receives statins) and the other is not exposed. [In reality, the two groups can never be identical; there will always be some random (chance) differences between them due to natural variability.] Consequently, if a clear difference is seen in the outcome measure (lung cancer risk or survival time), it should only be due to the exposure status, and not any other factor. This is a fundamental concept in medical research, and one that allows **causal** inferences to be made more reliably. An example is shown in Box 1.4.

In a randomised clinical trial, the process of randomisation aims to 'make everything the same', except the intervention given. The researcher randomly allocates the interventions (exposures) leading to two similar groups. Any

[#] Presented in Chapter 2

Box 1.4 Illustration of how differences between exposed and unexposed groups influence the effect of an exposure on an outcome measure

	Exposed: smokers	Unexposed: never-smokers
	N = 2500	N = 7500
Eat lots of fruit and vegetables	25%	60%
Had heart attack	10%	5%

Interest is only in examining the effect of smoking on the risk of a heart attack. The risk is twice as high among smokers than never-smokers, so we could conclude that smoking is associated with heart disease. But it is not possible to distinguish whether this difference (effect) could be due to:

- The difference in smoking status
- The difference in diet
- A combination of the two

differences in the outcome measure should only be due to the intervention, which is why clinical trials (and systematic reviews of them) usually provide the best level of evidence in medical research, and a causal relationship can often be determined. Published reports of all randomised studies contain a table confirming that baseline characteristics are similar between the trial groups.

In observational studies, however, the exposure cannot be randomly allocated by the research team. The researchers can only observe, not intervene, and it is likely for several differences to exist between the groups to be compared. The more differences there are, the more difficult it will be to conclude a causal link. The two main sources of these differences are **confounding** and **bias**. Confounding and bias might still be present to some small extent in a randomised clinical trial, but the purpose of randomisation is to minimise their effect.

Confounding and bias can each distort the results and therefore the conclusions (Box 1.5).

Some researchers consider confounding as a type of bias, because both have similar effects on the results. However, a key difference is that it is usually possible to measure confounding factors and therefore allow for them in the statistical analysis, but a factor associated with bias is often difficult or impossible to measure, and therefore it cannot be adjusted for in the same way as confounding. Confounding and bias could work together, or in opposite directions. It may not be possible to separate their effects reliably.

Researchers try to remove or minimise the effect of bias at the design stage or when conducting the study. The effect of some confounding factors can also be minimised at this stage (**matched case–control studies**, see Chapter 6, page 114).

Box 1.5 Confounding and bias

- **Confounding** represents the natural relationships between physical and biochemical characteristics, genetic make-up, and lifestyle and habits, which may affect how an individual responds to an exposure.
- It cannot be removed from a research study, but known confounding factors can be allowed for in a statistical analysis if they have been measured, or at the design stage (matched case–control studies).

- **Bias** is usually a design feature of a study that affects how participants are selected, treated, managed, or assessed.
- It often arises through the actions of the study participants and/or the research team.
- The effect of bias could be minimised or prevented by careful study design and conduct, but human nature makes this difficult.
- It is difficult, sometimes impossible, to allow for bias in a statistical analysis because it cannot be measured reliably.

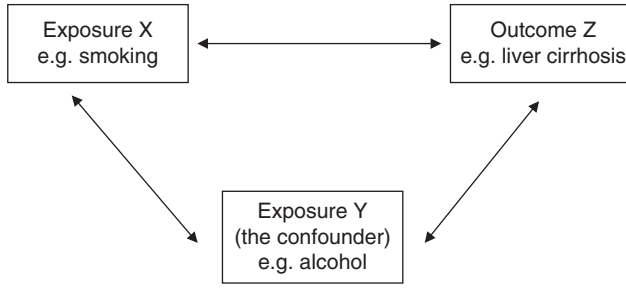
The confounding and bias factors themselves are relatively unimportant. What matters more is whether they greatly influence the study results:
- Make an effect appear spuriously, when in reality there is no association
- Overestimate the magnitude of an effect
- Underestimate the magnitude of an effect
- Hide a real effect

Confounding

A confounding factor is often another type of exposure, and to affect the study results, it must be associated with both the exposure and outcome of interest (Figure 1.1). The factor could be more common in either the exposed or unexposed groups.

Figure 1.2 shows a hypothetical example of how confounding can distort the results of a study. The primary interest is in whether smoking is associated with death from liver cirrhosis. In Figure 1.2a, if the death rates are simply compared between smokers and non-smokers, they appear to be higher among smokers (15 vs. 9 per 1000). It could be concluded that smokers have a higher risk, and this could be used as supporting evidence that smoking is a risk factor for cirrhosis. However, from Figure 1.2a, it is clear that smokers are more likely to be alcohol drinkers (66 vs. 34%), and it is already known that alcohol increases the risk of liver cirrhosis. Because the exposed (current smokers) and unexposed (never-smokers) groups have different alcohol consumption habits, they are not 'the same', and the difference in death rates could be due to smoking status, the difference in alcohol consumption, or a combination of the two.

Because drinking status has been measured for all participants, it is perhaps intuitive that to remove its confounding effect, the association between smoking and cirrhosis deaths can be examined *separately* for drinkers and non-drinkers.



A confounder Y (sometimes another type of exposure) can only distort the association between the exposure of interest X and outcome Z, if it is associated with both X and Z. In the example, people who smoke tend to drink alcohol, and people who drink alcohol have a higher chance of developing cirrhosis.

Figure 1.1 The effect of an exposure on an outcome, with a third factor, the confounder.

Interest is only in the association between smoking (the exposure) and death from liver cirrhosis (the outcome); i.e. whether people who smoke have a higher chance of dying from cirrhosis than people who have never smoked.

In this hypothetical study, there are 1000 smokers and 1000 never-smokers.

(a)

	Current smokers	Never-smokers
Death rate from liver cirrhosis	15 per 1000	9 per 1000
% who drink alcohol	66	34

(b)

	Current smokers		Never-smokers	
	No. deaths/no. of men	Death rate per 1000 (A)	No. deaths/no. of men	Death rate per 1000 (B)
All	15/1000	15	9/1000	9
Non-drinkers	1/340	3	2/660	3
Drinkers	14/660	21	7/340	21

Figure 1.2 Hypothetical example of how a confounder can distort the results when examining the effect of an exposure on an outcome and how it can be allowed for.

This is shown in Figure 1.2b. By comparing the death rates between smokers and never-smokers only among non-drinkers, alcohol cannot have any confounding effect, because the two exposure groups have been 'made the same' in terms of alcohol consumption. The death rates are found to be identical,

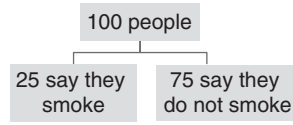
and the conclusion is reached that smoking is not associated with cirrhosis in this group. A similar finding is made among drinkers only, where, although the death rates are higher than those in non-drinkers (as expected), they are identical between smokers and never-smokers. The effect of confounding has been to create an association when really there was none. Analysing the data in this way (called a **stratified analysis**) is the simplest way to **allow** or **adjust for** a confounding factor. In practice, there are more efficient and sophisticated statistical methods to adjust for confounders (**regression analyses**; Chapter 4). If there is uncertainty over the relationship between the confounder and either the exposure or the outcome, it is worth taking it into account as a **potential confounding factor**.

A factor should not be considered a confounder if it lies on the same biological (causal) pathway between the exposure and an outcome [2]. For example, if looking at the effect of a high-fat diet on the risk of heart disease, high cholesterol is a consequence of the diet, and it can also lead to heart disease. Therefore, cholesterol would not be a confounder because it must, by definition, be causally associated with both exposure and outcome, and its effect should not (or cannot) be removed.

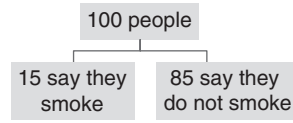
Bias

A **bias** occurs where the actions of participants or researchers produce a value of the study outcome measure that is *systematically* under- or over-reported in one group compared with another (i.e. it works in one direction). Figure 1.3 is a simple illustration. In the middle figure, only people who smoke lie about (misreport) their smoking status, and the effect of this is to **bias** the study result (in this case the prevalence of smoking). If, however, the number of non-smokers who lie about their smoking status is similar to that in smokers, even though there are lots of people who misreport their habits, the study result itself is not biased. But non-smokers rarely report themselves as smokers. It is important to focus on the bias in the *result* rather than the factor creating the bias.

Unlike confounding (where in the example above it was simple to obtain the alcohol status of the study subjects and, therefore, allow for it when examining the effect of smoking on liver cirrhosis), it is difficult to measure bias, because it would require the participants to admit whether or not they are lying, which, of course, would not happen. Researchers attempt to minimise bias at the design stage. In the example in Figure 1.3, estimating smoking prevalence could be assessed using biochemical confirmation of smoking status using nicotine or cotinine in the saliva of the participants, where high levels are indicative of being a smoker. However, even this is not perfect, because a light smoker could have low concentrations that overlap with non-smokers, and non-smokers heavily exposed to environmental tobacco smoke could have levels that overlap with some smokers. Many other biases are similarly difficult or impossible to measure.

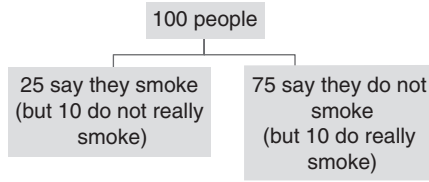


If no one has lied, then the true (and observed) prevalence of smoking is 25%. (25/100)



But if 10 of the 25 smokers lie and none of the non-smokers have lied, these 10 would be counted in the non-smoking group. The observed smoking prevalence would then be 15% (compared with the true value of 25%).

The study result would be **biased,** and **under-estimated**.



But if 10 of the smokers lie and 10 of the non-smokers have also lied, then although there are those who misreport their smoking status in both groups, the observed smoking prevalence would be 25% (the same as the true value).

The study result is not biased.

Figure 1.3 Illustration of bias, using an example in which the aim is to estimate the proportion of people who smoke, based on self-reported measures.

There are several types of biases (Box 1.6), and they can arise from something either the researcher or study participant has done [4]. To determine whether bias exists, the following questions should be considered:

- Was there anything unusual in how the participants were selected for the study?
- Were some participants managed, assessed, or examined differently from others?
- Is it plausible that certain participants could misreport, or under- or over-report, their responses to a questionnaire and hence distort the results?

1.3 Types of observational studies

Studies are conducted among two different types of participants:

1. **Population**: Participants are approached from the general population. They may or may not have the disorder of interest. Researchers sometimes

Box 1.6 Common types of potential biases

- **Selection bias**: The participants chosen for the study are not representative of the population of interest. An example is the healthy worker effect, where disease rates are lower in the study group than in the general population.
- **Response/responder (or non-response) bias**: People who agree to take part in a study have different characteristics from those who do not, and this distorts the results when making conclusions about the whole population.
- **Recall bias**: People with disease are often better at remembering past details (including past exposures) about their life than people without disease.
- **Withdrawal bias**: Participants who decide to discontinue with a study have different characteristics from those who continue, and this can distort the results because follow-up data (e.g. outcome measures) will be missing for some participants.
- **Assessment bias**: Different groups of participants are managed using different assessments or at different times according to their characteristics, exposure status, or health outcome.
- **Measurement bias**: Measuring exposures is performed differently for people with different health outcomes.
- **Observer or interviewer bias**: If an interviewer is aware of the participant's health (or other) status, this may influence the questions asked, or how they are asked, which consequently affects the response.

use the word **healthy** individual or **control** when describing some study participants, but this usually only means that the participants do not have the disorder of interest. They may have other disorders. Better terms could be **affected** and **unaffected**.

2. **Patients**: Only people who have already been diagnosed with a specific disorder are recruited to a study.

The study objectives are usually quite different for each of these two types. For studies of the **population**, interest is often in risk factors that lead to the occurrence of a disorder, but for **patient** studies interest could be in how an existing disorder develops including the management of it. Both can be used when describing characteristics of a group(s).

A variety of study designs can be used to examine associations, risk factors, and interventions.

- A **cross-sectional survey**: face-to-face interviews with participants or collecting self-completed participant surveys.
- A **(retrospective) case–control study**: people with and without a disorder of interest are identified and asked about their past habits, possibly also obtaining data from their medical records.
- A **prospective cohort study**: people without the disorder of interest are identified, baseline characteristics are measured, and participants are followed up for a period of time (several months or years) during which specific data is collected regularly.

- A **retrospective cohort study** is essentially a prospective cohort study that has already been conducted.
- **Longitudinal study**: a prospective cohort study in which exposures and often outcomes are measured repeatedly during follow-up.
- Studies based on routinely collected data: these could come from **regional or national registries or databases** (e.g. cancer or death notification systems) and contain a few key factors on each individual (e.g. age, sex, city of residence), as well as the disease status. Many such databases have adequate or good data quality processes in place, but a common limitation is that potential confounding factors are unavailable.

These terms for types of study designs should not be regarded as fixed. There may be occasions when one type could be used synonymously with another, a design is nested within another, or there are variations on a specific design. For example:

- There are nested case–control studies, which involve selecting and only analysing cases and controls (individuals with and without a disorder of interest) from a cohort study.
- Cases and controls could provide information about their current or past characteristics, but they might also be followed up for a certain length of time for other outcome measures, so these data are collected prospectively (similar to a prospective study).

Researchers simply need to be clear where the participants for a particular study have come from and how data are collected from or about them.

Large-scale studies could be preceded by a **pilot study** to examine the likely recruitment rate and how data are to be collected. Problems that arise can be dealt with before launching the full study. Pilot studies should have few participants and have a short duration.

An **ecological study** is one in which the unit of interest is a group of people, not an individual. For example, the relationship between income and risk of heart disease could be examined by using the average income from 20 countries and the corresponding rates of heart disease in each country, and then examine the correlation. However, such studies can often only provide a crude measure of association because potential unmeasured confounding factors could explain the effect (ecological fallacy); confounding is best dealt with at an individual level. The findings in ecological studies can therefore be inconsistent with those based on individuals.

A common but special type of observational study is a **qualitative research study**. This is usually based on relatively few participants (often <50). Although a structured questionnaire could be used to ascertain some information about the participants, the main source of data is by face-to-face or telephone interviews, with largely open-ended questions to find out about their characteristics, lifestyle habits, opinions, or experiences (other study types almost always use structured questions). The interviews are usually recorded, allowing researchers to play back the recording later and code the responses in a way that can be interpreted and summarised. The findings are often descriptive, and the

data produced cannot be readily quantified, and therefore not analysed using statistical methods covered in Chapter 4. For these reasons, they are not discussed in detail in this book, but they are well described elsewhere [5, 6]. A qualitative study can be used:

- As a precursor to the study designs mentioned previously in order to better design the questionnaires for a larger and more structured study (i.e. how to measure factors, exposures and outcomes), or to obtain an initial understanding of the research question
- To attempt clarification of some of the findings of studies, or a deeper understanding of them, especially if they are unexpected

Other types of observational studies include **case or case series reports**, which are based on unusual or sporadic occurrences found by a health professional, often during clinical practice [7]. They may provide interesting findings, but no firm conclusions should be made from these. They usually lead to better designed studies.

1.4 Strengths and limitations of the different types of study designs

There are ways of assessing the reliability of evidence from a particular study, such as the Grading of Recommendations Assessment, Development and Evaluation (GRADE) [8]. The most reliable type of study is, in order (generally, but there are exceptions):

- Systematic review of randomised clinical trials
- Individual randomised clinical trial
- Systematic review of observational studies
- Individual prospective cohort study
- Individual case–control study
- Individual cross-sectional study
- Hospital audits
- Case reviews

When examining risk factors and causality, there are many situations where a randomised trial cannot be conducted. For example, the best way to determine that smoking is a cause of cancer is to randomly allocate never-smokers to either take up smoking regularly for several years or remain non-smokers, then follow them up and compare the proportions that develop cancer between the two groups. This would clearly be unethical, so the only way to examine this risk factor is by using a cohort or case-control. Also, studies of aspects such as patient satisfaction and quality of care are generally descriptive, in which case a cross-sectional survey is the preferred method, because the primary purpose is not to look at the effect of an exposure on an outcome.

There can be some overlap between the different designs and more than one may be appropriate for a particular research question. A key distinction between them is **time**. Data can be collected retrospectively, from past participants/patients (hospital audits and case–control studies) or at one point in time, usually the present (cross-sectional studies). Alternatively, it can be collected prospectively, after entry to the study over a few months or several

years, from newly identified participants/patients (cohort studies). It is also possible to conduct a retrospective cohort study, where data on exposures and outcomes have already been obtained, but there has still been a sufficient length of time between them. Such a study can be considered as a prospective study that has already been conducted.

Box 1.7 shows several strengths and limitations of different study designs, for consideration when choosing one over another. A cohort study is generally more reliable than a case–control study, but for uncommon disorder a cohort study conducted for several years that can only ascertain 50 cases is much less preferred than a case–control study in which 500 cases can be found quicker.

There is no such thing as the perfect study, regardless of how well it is designed. With hindsight, all investigators can identify ways in which their study could have been improved, having encountered problems and issue not expected at the start of the study that are associated with the design itself, data collection, and statistical analyses.

Observational studies for examining interventions

Observational studies have often been used to examine the efficacy or safety of an intervention [9], and they are a major feature of **comparative effectiveness research** [10–12]. However, there are design limitations that can produce a spurious or overestimated treatment effect (due to bias or confounding), and a randomised clinical trial is almost always preferred in this situation. Findings from observational studies can be consistent with those from randomised clinical trials (the gold standard for evaluating interventions). For example, a systematic review of 20 observational studies indicated that giving the influenza vaccine to the elderly could halve the risk of developing respiratory and flu-like symptoms [13], and the same effect was found in a large double-blind clinical trial [14].

A potential strength of observational studies, compared to randomised trials, is that they can provide supporting evidence for an intervention because the participants might be more representative of the target population (people who participate in clinical trials can sometimes be a self-selected group with different characteristics from those who decline to take part). Also, the study size is often larger than a clinical trial, making it easier to examine side effects, particularly those that are relatively uncommon.

However, there are situations where a treatment benefit has been found in observational studies, but not in a randomised clinical trial, or over-estimated or the opposite conclusion made. An example of the latter is where observational studies indicated that people with a high β-carotene intake (lots of fruit and vegetables) had a lower risk of cardiovascular death than those with a low intake (31% reduction in risk) [15], but randomised trials showed that a high intake might increase the risk by 12% [15].

1.5 Key design features

When conducting an observational study, there are several important design features to consider (covered in more detail in Chapters 5–8)

- Which study participants should be included (eligibility criteria)?
- Where will they come from (sampling frame)?

Box 1.7 Some strengths and limitations of the main observational study designs

Study design	Strengths	Limitations
Cross-sectional survey (retrospective).	• Can be inexpensive and quick to conduct • Can estimate prevalence of a disorder or event. • Usually no concept of loss to follow-up (i.e. participants who drop out of the study). • Can examine several exposures and several outcomes.	• Response rates may be low, leading to problems selection or responder bias. • Recall bias can be a problem if participants are asked about past information. • If examining an intervention, it may not be possible to ascertain from records why some patients received the intervention and others did not. • Can only examine association between an exposure and outcome (not causality).
Case–control study (retrospective).	• Can be relatively inexpensive and quick to conduct. • Suitable for studying rare disorders or events. • Usually no concept of loss to follow-up. • Can examine several exposures in relation to a single disorder.	• Not suitable for studying uncommon exposures. • Could be affected by recall and selection biases. • Not possible to determine prevalence or incidence of disease. • It can sometimes be difficult to establish whether the exposure came before the outcome (required for causality). • Care is required when selecting cases and controls (i.e. that they both come from the target population of interest).
Prospective cohort study.	• Prospective follow-up means that exposures and outcome measures should be easier to record. • Can estimate the incidence of a disease. • Possible to look at how exposure changes over time. • Less chance of selection or recall bias. • It is known that the exposure came before the outcome (required for causality). • Can examine the natural history of a disorder. • Can examine several disorders in relation to a single exposure. • Can examine several exposures in relation to a single outcome.	• Can be expensive, both in terms of money and staff resources. • Can take a long time before final results are available. • Not suitable for investigating rare disorders or events. • If the exposure status changes over time, this needs to be ascertained, and the statistical analyses are more complex. • Potentially many participants could drop out (especially with long follow-up), so there would be no measure of the study endpoints on these. • If the characteristics of those who drop out are very different between the exposed and unexposed groups, this could create bias. • Could be affected by non-response and time-related biases.

Retrospective cohorts have similar features to prospective ones, except they are relatively inexpensive but can be affected by missing data.

- What will be done to them, and how often (collecting data)?
- How will the exposures, confounding factors, and outcome measures be measured?
- How will potential confounding and bias be minimised or addressed?

Eligibility criteria and recruitment

In many observational studies, the study population should be defined by a set of **inclusion and exclusion criteria**. They specify which participants are recruited or whose data to include (if the study is based on patient medical records or established databases). Each subject has to meet the criteria before being included in the study, though there may be acceptable small deviations. The criteria depend on the research objectives, and may include an age range and the ability to provide informed consent. Eligibility criteria should have unambiguous definitions to make recruiting participants easier. Some studies, such as those based only on patient records, might have few or no criteria, because they are based on every patient with a certain disorder.

Determining the eligibility criteria necessitates balancing the advantages and disadvantages of having a highly selected group against those associated with including a wide variety of participants. Having many narrow criteria (e.g. age range of 30–35) produces a group in which there should be relatively little variability ('make everything the same'), and it is easier to find associations especially if the effect is small or moderate. However, the study results may not be easily generalisable. A study with few criteria that are wide (e.g. age ≥18) will have a more general application, but the amount of variability could make it more difficult to detect an association, and sometimes only large effects can be found easily.

Not everyone who is eligible for a particular study will agree to participate, and the higher the acceptance rate, the better. However, if, for example, more than 40% decline to take part, it can be useful to attempt to obtain some information from these participants (e.g. characteristics such as age, gender, and some measure of the exposure factor), which can be used to compare with those who agreed to participate. The study results could be biased if those who refused and those who participated are very different. Reasons for non-participation could also be used to redesign parts of the study while it is being conducted to improve uptake.

Encouraging patients to take part in a clinical trial of a treatment may be easier than an observational study, because they see a potential personal benefit, assuming the treatment is effective. However, this is not the case for observational studies, so maximising uptake is worthwhile (Box 1.8).

Sampling frame

A key design feature of all observational studies is the **sampling frame**. This is a list of people from which the target group of interest will be identified. It is essentially the starting point of a study. There are many examples of sampling frames; they can be local to the researchers, where access is easy (e.g.

Box 1.8 Possible ways of encouraging people to take part in an observational study (if it involves recruiting people)

- Make clear what the potential benefits are to society and possibly themselves (e.g. identifying new lifestyle risk factors that individuals can modify after the study results are available)
- Minimise inconvenience associated with collecting data and measuring exposures and outcomes
- Provide costs to cover travel and subsistence if the study involves attending clinics for assessments
- Provide information about or discuss possible anxieties people may have about health issues to be raised by the study

people registered at a single physician's practice or listed outpatients in a single hospital respiratory clinic); **regional** (e.g. all adults listed on the census or registry in a town or geographical region or found using telephone directories); or **national** (e.g. all registered general practitioners in England, access to adults listed on all censuses or registries in a country, or a register for a specific disorder).

The choice of sampling frame (local, regional, or national) will depend on the research question and how representative the research results need to be. Examples of sampling frames and research objectives are given in chapters 5–8, and they should allow the conclusions of the study to be generalisable to the wider population of interest. For example, a study of factors that influence the severity of chronic obstructive lung disease (COPD) could be conducted in a respiratory clinic at a hospital, but this would exclude COPD cases who only see their family physician and may therefore have milder disease. Also, researchers use local sampling frames because they have limited staff or funds.

Once the sampling frame has been determined, the next question is whether to include everyone within it or a **random sample**. This choice, again, is influenced by costs and feasibility. If a random sample is used, the process should be explicitly described, and there are various methods for appropriately selecting participants at random, to help to ensure that some participants are not chosen in a way that could produce a bias [16].

Collecting data

All observational studies involve collecting data, which may or may not require direct input from the study participants. There are several ways in which this could be done (Box 1.9). Different sources of data have different attributes, such as quality, validity, reliability, and measurement error [17]. Obtaining information directly from participants will be a choice between interviews or self-completed questionnaires, and each has strengths and limitations (Box 1.10). Missing data is a major problem that usually cannot be

Box 1.9 Sources of data within observational studies

- Face-to-face or telephone interviews directly with study participants.
- Self-completed questionnaires (handed or posted back to the researchers), including self-completed diaries.
- Face-to-face interviews or questionnaires completed by a proxy for the study participant (e.g. close relative or friend).
- Biological samples (e.g. blood, urine, saliva, or tissue).
- Imaging tests (e.g. X-ray, CT, or MRI scan) or clinical examinations.
- Environmental measures (e.g. air pollution, quality of drinking water).
- Health records from family or primary care clinics or hospitals.
- Local, regional, or national registries/database that routinely record population data on, for example, deaths and cause of death, occurrence of cancer, occurrence of specific disorders, or hospital admissions. Study participants would need to be **linked** to these databases using personal identifiers.

These methods could be used on their own or in combination for a particular study.

Box 1.10 Strengths and limitations of obtaining information from study participants using either interview or self-completed questionnaires

Interview	Self-completed questionnaire
Requires dedicated staff to meet with each participant at their home or a research site, or to interview by telephone (can therefore be expensive)	Can be sent out and received by post, allowing a wider coverage of the sampling frame (can therefore be relatively inexpensive)
Direct contact encourages participants to respond and to complete most/all questions	Can be affected by moderate to high non-response rates and missing data for several questions
Complex questions misunderstood by the participant could be clarified by the interviewer	Complex questions can be difficult to interpret. This can be helped using clear instructions on how to complete these questionnaire fields
Interviewer may have limited time to spend with each participant, so the questionnaire should be relatively short	Participant can complete the questionnaire in their own time and therefore more questions could be included
Useful when responses to some questions need probing or further clarification	Useful for questionnaires containing sensitive questions, including those that require anonymity
Interviewers could influence (bias) the responses, particularly if they are aware of the participant's exposure or disease status and are aware of the study objectives	The researchers cannot directly influence the responses

overcome with most retrospective studies, but attempts could be made to minimise this in prospective studies by using good data collection systems (e.g. simple/short questionnaires; see Box 5.6). Also, regular (e.g. yearly) general updates of the study to all participants could keep them interested, minimising dropouts (and therefore withdrawal bias).

A prospective cohort study is generally the only type of observational study that usually involves collecting data directly from participants over time. This is often done using questionnaires, but some studies might involve physical and clinical examinations, and collection of biological samples or imaging tests. A **schedule of assessments** must be drawn up, stating explicitly when contact is to be made with the participants and how data are to be collected at each time point. A vital part of follow-up is the evaluation of the outcome measure, which may be done through regular reviews of clinic records or by linking the study participants to regional or national registries that routinely collect information on disease occurrence or deaths.

With advances in information technology, more people now have personal computers, mobile telephones, or smartphones, and observational studies will probably make use of these. Study participants could complete questionnaires online (rather than face-to-face with a researcher) or by using a Personal Digital Assistant provided by the researchers specifically for the study. These approaches have the potential advantages of targeting a wider and more generalisable population (larger sampling frame), increasing response rates, and decreasing study costs (reduced central data management because there will be much less data to be entered manually). Key considerations in dealing with these advances will be that such studies may only be tenable in countries where many/most people have access to personal computers, and there may be issues over representativeness (whether characteristics differ significantly between people who do and do not have a computer) and security (accessing a central research database and exchanging sensitive patient confidential data that require the anonymity of the participant.)

Studies using only patient medical records

Patient medical records (from hospitals, family physicians or registries) are sometimes used as the only source of data for observational studies. This is particularly useful when there is limited time, because the data already exist and nothing is required from patients directly. However, such studies are almost always based on patients seen in routine practice, so researchers can only use data that have already been collected. Data can be clinical, blood or imaging measurements, pathology results, or standard characteristics such as sex, age, ethnic origin, and disease status.

Observational studies that use stored records usually consist of people who already have a disorder or are seeking professional health care, instead of individuals from a general 'healthy' population. Therefore, a common objective is to provide simple descriptive statistics on a defined group of patients, or to examine an intervention or a care pathway. They can also be used to investigate associations or prognostic markers.

Box 1.11 Key characteristics of studies according to method of data collection

Manual extraction	Electronic data systems
The person extracting the data may form a reasonable view of general aspects of the data such as quality	Can provide a large number of patients
Potentially important factors/ measurements that had not been originally planned could be identified and collected during data extraction	Selecting patients, using the predefined eligibility criteria, could be more accurate
Given limited resources, only one or few searches of the records could be made	Multiple searches of the database can be made easily

Choosing which patients to include involves specifying clear and simple selection criteria, and a time frame (e.g. all newly diagnosed cases of thyroid cancer between 1995 and 2012). Too many selection criteria could limit the number of patient records for the study. However, the main problems are missing data and data quality (inconsistencies cannot usually be clarified nor errors corrected).

Older data, from patients seen many years ago, are more likely to be stored on paper (in the clinic, or may have to be retrieved from an off-site archive). Extracting such data can be laborious, requiring staff to examine each patient record and manually transfer the factors (variables) of interest onto a study data sheet. All of the factors should be pre-specified, to avoid staff having to conduct repeated searches and extractions of the same patient records. Many health service providers now use electronic patient data systems, and it may appear relatively simple to download a set of specific variables for a defined group of patients. However, many clinical IT systems were not set up for research purposes, so downloading data may require a dedicated programmer. IT support may also be required to collect and merge data from several clinics, especially if they use different software systems. Some key characteristics of manually extracting data or obtaining it from electronic databases are shown in Box 1.11.

Clear definition of the exposures and outcome measures

The key factors (variables) of interest in a study, especially those that involve examining the effects of exposures on outcome measures, require clear definitions as do all potential confounding factors. Having well-defined endpoints and objectives will facilitate:

- Conduct of the study (e.g. the researchers are focussed)
- Decisions on what data (information) to collect and how to do this

- Analysis of the data
- Interpretation of the results
- Writing of the final report for publication

It will also help to reduce significant criticism of the paper when submitted for publication in a peer-reviewed journal, acknowledging that there may not be perfect (standard) definitions of either the exposures or the outcomes, and some may disagree with a chosen definition. The key factors should be measured objectively, rather than subjectively, where possible (e.g. measuring carbon monoxide in exhaled breath versus self-report in a study of smokers who quit).

Many exposures may initially appear easy to define, but on closer inspection they often have several descriptions. For example, if examining the effect of alcohol consumption habits on emergency hospital admissions for physical injuries, 'alcohol drinking' could be measured as any of the following:

- Someone who regularly drinks alcohol (i.e. at least once per week)
- Number of units drunk in previous week
- Number of units drunk over a typical month

Ideally, outcomes (such as disorders) should be measured using standard and generally accepted methods, (e.g. histopathology for cancer) or established diagnostic tools (e.g. *Diagnostic and Statistical Manual of Mental Disorders* for psychological disorders).

In this book, a variety of exposures and outcomes are used as examples. It is useful to realise that they can be analysed and interpreted in a similar way:

- Clinical features or characteristics
- Environmental exposures
- Lifestyle habits or characteristics
- Imaging marker
- Biomarker
- Intervention
- Perceived experiences and measures of satisfaction

Consideration of confounding and bias

Information on known and potentially important confounding factors should be collected as part of the study. In a study of long duration, it might be possible to add 'new' confounding factors to the **case report forms** at a later date (see Chapter 11, page 226).

Taking account of the common types of bias (Box 1.6) can usually help to design the study to avoid or minimise the effects, though this is sometimes difficult:

- Careful selection of study participants, without choosing them on the basis of factors of interest, can minimise selection bias, for example, not trying to recruit heavier smokers for a study of the association between smoking and a disorder.
- Where possible, objective measures of exposures and outcomes should be used.
- Where possible, have an independent review of the outcome measures, ideally where the reviewer is blind to (i.e. unaware of) the exposure status of the participants.

1.6 Interpreting and reporting the results and implication for public health or clinical practice

A major task for the investigators is to analyse and interpret the findings, and communicate the results through conferences and journal articles. The following structure will be used in Chapters 5–7, in the sections entitled 'Analysing data and interpreting results'.

What are the main results?

It is important to focus on the main result(s) in the context of the study objectives. Researchers should examine this first and ensure they understand this (quantitatively), including clinical importance and any implications.

What could the true effect be, given that the study was conducted on a sample of people?

After the main result(s) have been interpreted, it is necessary to examine **95% confidence intervals** (90 or 99% are alternatives), because these will provide a likely range of the *true* effect (see Chapter 3).

Could the observed result be a chance finding in this particular study?

Statistical significance (p-values) is a useful part of the analyses, but it is important to understand them fully, including what influences the size of a p-value (see Chapter 3).

How good is the evidence?

Considering major strengths and limitations of a study and the study findings is essential, including whether the findings and conclusions are generalisable. Also, whether the interpretation of the findings is likely to have been influenced by bias or confounding. If there were significant confounding factors or bias, is the main result unreliable? No individual study should change practice. There should always be corroborating evidence from at least one other study, and this can include:

- Other similar studies (same exposure and similar population)
- Studies of the same exposure in different populations
- Studies investigating the biological plausibility, including laboratory evidence (i.e. whether the association makes sense)

Finally, researchers should always attempt to suggest what should happen next, and discuss how the main findings and conclusions of their study should be used, for example, to change clinical or public health practice or to make recommendations for further research.

1.7 Translational research

It is becoming increasingly common to collect biological specimens as part of the main study, to be stored centrally in a laboratory for either pre-specified analyses to be performed at the end of the study, or for future as yet unspecified

Box 1.12 Some key considerations for studies examining biomarkers

- A central laboratory should be used, unless the marker is well-established and commercial assays are available.
- Good systems must be in place for collecting, processing, and shipping biological samples, for all recruiting centres and the central laboratory.
- If the samples are obtained from many study participants, across many centres (including serial samples over time from the same participant), there should be (electronic) systems for tracking the samples from the centre to the central laboratory.
- Secure and well-maintained storage of samples should be set up, e.g. fridges or freezers, with proper labelling and coding of each sample allowing it to be retrieved easily, and matched (anonymously) to the correct study participant. Electronic barcode readers could be useful when there are many samples, because this reduces human coding errors. Also, the storage facilities (e.g. freezer/fridge temperature) should be monitored continuously.
- Quality control processes should be in place continuously (this might include repeated measurements of the same samples).
- The laboratory assay or technique to measure the marker should have been validated, and error/failure rates examined, including measurement error.
- Semi-automated or manual assessment (scoring) of samples needs clear specification, and ideally. each sample (or a random subset) needs to be assessed by two independent people.
- Should tissue be collected from people with or without a disorder of interest, or disease/damaged and healthy tissue from the same persons.

analyses. This will involve the creation and maintenance of a **biobank**. The samples are usually blood, saliva, or urine, but may also include tissue samples (e.g. cancerous tissue removed from affected patients). The analyses involve measuring biomarkers, which could be chemical or biological (e.g. genetic or protein markers), or imaging markers, and these are correlated with clinical outcomes from the main study, such as the risk of disease, disease severity, or mortality. Many biological or imaging factors can be analysed like traditional (external) exposures.

One of the main purposes of these analyses is to examine the prognostic value of a marker, that is, how well it **predicts** a clinical outcome (Chapter 8). A key issue is the reliability of a marker and that it has been properly validated (i.e. it measures what it is meant to measure). Box 1.12 shows important features of biomarker studies.

Not all studies will benefit from having a translational study component, and indeed the collection and storage of biological samples could sometimes be a hindrance to the main study, by adding significant extra financial costs and time to collect the samples. Researchers should decide whether translational research is essential for their study.

Biological samples could be used to examine:

- The relationship between the biomarker and an exposure (e.g. lifestyle, environmental, or clinical characteristics)
- The relationship between the biomarker and a disorder (or other event)
- Methods for detecting or diagnosing disease
- Methods for detecting infectious or microbial agents
- Biological mechanisms and processes
- Surveillance (or monitoring) within a population

This field of work can be called **molecular epidemiology** [18].

When using biological samples, it is important that enough material is collected for the study objectives, and this should be agreed with the laboratory staff (e.g. minimum amount of blood). Also, how the samples are to be handled, for example, stored at room temperature or in fridge or freezer, and whether samples need to be posted immediately by courier or in batches.

1.8 Key points

- Natural variability or variation underpins aspects of study design and analysis.
- Observational studies can be used to describe the characteristics of a single group of people.
- Another major purpose is to examine the effect of an exposure on an outcome.
- To do this reliably, we need to make the exposed group 'the same' as the unexposed group (except the exposure factor of interest).
- Confounding and bias are the most common reasons why the exposure groups are not 'the same'.
- There are several types of observational studies, each with strengths and limitations: cross-sectional, case–control, and retrospective or prospective cohort.
- All studies need clear definitions of factors, exposures, and outcomes.
- Reliable processes should be in place for dealing with biological specimens for translational research.

References

1. Silman AJ, Macfarlane GJ. *Epidemiological Studies: A Practical Guide*. Cambridge University Press. Second Edition (2002).
2. Rothman KJ. *Epidemiology: An Introduction*. Oxford University Press. First Edition (2002).
3. Barker DJP, Rose G. *Epidemiology in Medical Practice*. Churchill Livingstone. Fourth Edition 1990.
4. Grimes DA, Schulz KF. Bias and causal associations in observational studies. *Lancet* 2002;359:248–52.
5. Pope C, Mays N. *Qualitative Research in Health Care*. Wiley-Blackwell. Third Edition. (2006).
6. Silverman D. *Doing Qualitative Research: A Practical Handbook*. SAGE Publications. Fourth Edition (2013).

7. Grimes DA, Schulz KF. Descriptive studies: what they can and cannot do. *Lancet* 2002;359: 145–9
8. http://www.gradeworkinggroup.org/index.htm. Accessed 14 May 2014.
9. MacMahon S, Collins R. Reliable assessment of the effects of treatment on mortality and major morbidity, II: observational studies. *Lancet* 2001;357:455–62
10. Sox HC, Goodman SN. The methods of comparative effectiveness research. *Annu Rev Public Health* 2012;33:425–45.
11. Armstrong K. Methods in comparative effectiveness research. *J Clin Oncol* 2012;30:4208–14
12. Korn EL, Freidlin B. Methodology for comparative effectiveness research: potential and limitations. *J Clin Oncol* 2012;30:4185–7.
13. Gross PA, Hermogenes H, Sacks HS, Lau J, Levandowski RA. The efficacy of influenza vaccine in elderly persons. *Ann Intern Med* 1995;123:518–27.
14. Govaert TME, Thijs MCN, Masurel N, Sprenger MJW, Dinant GJ, Knottnerus JA. The efficacy of influenza vaccination in elderly individuals. *JAMA* 1994;272(21):1661–5
15. Egger M, Schneider M, Davey SG. Spurious precision? Meta-analysis of observational studies. *BMJ*. 1998;316(7125):140–4
16. Moser C, Kalton G. *Survey Methods in Social Investigation*. Dartmouth Publishing Co. Second Edition (1985)
17. dos Santos Silva I. Measurement of exposures and outcomes. In: *Cancer Epidemiology: Principles and Methods*. IARC Press (1999). http://www.iarc.fr/en/publications/pdfs-online/epi/cancerepi/. Accessed 14 May 2014.
18. Rothman N, Hainaut P, Schulte P, Smith M, Boffetta P, Perera F, eds. *Molecular Epidemiology: Principles and Practices*. International Agency for Research on Cancer (IARC) (2011).