12 Study design I

Study design is vitally important as poorly designed studies may give misleading results. Large amounts of data from a poor study will not compensate for problems in its design. In this chapter and in Chapter 13 we discuss some of the main aspects of study design. In Chapters 14–16 we discuss specific types of study: clinical trials, cohort studies and case–control studies.

The aims of any study should be clearly stated at the outset. We may wish to estimate a parameter in the population (such as the risk of some event (Chapter 15)), to consider associations between a particular aetiological factor and an outcome of interest, or to evaluate the effect of an intervention (such as a new treatment). There may be a number of possible designs for any such study. The ultimate choice of design will depend not only on the aims, but on the resources available and ethical considerations (see Table 12.1).

Experimental or observational studies

- **Experimental** studies involve the investigator intervening in some way to affect the outcome. The clinical trial (Chapter 14) is an example of an experimental study in which the investigator introduces some form of treatment. Other examples include animal studies or laboratory studies that are carried out under experimental conditions. Experimental studies provide the most convincing evidence for any hypothesis as it is generally possible to control for factors that may affect the outcome. However, these studies are not always feasible or, if they involve humans or animals, may be unethical.
- **Observational** studies, for example cohort (Chapter 15) or case–control (Chapter 16) studies, are those in which the investigator does nothing to affect the outcome, but simply observes what happens. These studies may provide poorer information than experimental studies because it is often impossible to control for all factors that affect the outcome. However, in some situations, they may be the only types of study that are helpful or possible. **Epidemiological studies**, which assess the relationship between factors of interest and disease in the population, are observational.

Table 12.1 Study designs.

Type of study	Timing	Form	Action in past time	Action in present time (starting point)	Action in future time	Typical uses
Cross-sectional	Cross-sectional	Observational		Collect all information		• Prevalence estimates • Reference ranges and diagnostic tests • Current health status of a group
Repeated cross-sectional	Cross-sectional	Observational		Collect all information	Collect all information; Collect all information	• Changes over time
Cohort (Chapter 15)	Longitudinal (prospective)	Observational		Define cohort and assess risk factors	follow → Observe outcomes	• Prognosis and natural history (what will happen to someone with disease) • Aetiology
Case–control (Chapter 16)	Longitudinal (retrospective)	Observational	Assess risk factors	← trace Define cases and controls (i.e. outcome)		• Aetiology (particularly for rare diseases)
Experiment	Longitudinal (prospective)	Experimental		Apply intervention	follow → Observe outcomes	• Clinical trial to assess therapy (Chapter 14) • Trial to assess preventative measure, e.g. large scale vaccine trial • Laboratory experiment

Assessing causality in observational studies

Although the most convincing evidence for the causal role of a factor in disease usually comes from experimental studies, information from observational studies may be used provided it meets a number of criteria. The most well known criteria for assessing causation were proposed by Hill[1].

- The cause must precede the effect.
- The association should be plausible, i.e. the results should be biologically sensible.
- There should be consistent results from a number of studies.
- The association between the cause and the effect should be strong.
- There should be a dose–response relationship with the effect, i.e. higher levels of the effect should lead to more severe disease or more rapid disease onset.
- Removing the factor of interest should reduce the risk of disease.

Cross-sectional or longitudinal studies

- **Cross-sectional** studies are carried out at a single point in time. Examples include surveys and censuses of the population. They are particularly suitable for estimating the **point prevalence** of a condition in the population.

$$\text{Point prevalence} = \frac{\text{Number with the disease at a single time point}}{\text{Total number studied at the same time point}}$$

As we do not know when the events occurred prior to the study, we can only say that there is an association between the factor of interest and disease, and not that the factor is likely to have *caused* disease. Furthermore, we cannot estimate the **incidence** of the disease, i.e. the rate of new events in a particular period (Chapter 31). In addition, because cross-sectional studies are only carried out at one point in time, we cannot consider trends over time. However, these studies are generally quick and cheap to perform.

- **Repeated cross-sectional** studies may be carried out at different time points to assess trends over time. However, as these studies involve different groups of individuals at each time point, it can be difficult to assess whether apparent changes over time simply reflect differences in the groups of individuals studied.
- **Longitudinal studies** follow a sample of individuals over time. They are usually **prospective** in that individuals are followed forwards from some point in time (Chapter 15). Sometimes **retrospective** studies, in which individuals are selected and factors that have occurred in their past are identified (Chapter 16), are also perceived as longitudinal. Longitudinal studies generally take longer to carry out than cross-sectional studies, thus requiring more resources, and, if they rely on patient memory or medical records, may be subject to bias (explained at the end of this chapter).

Experimental studies are generally prospective as they consider the impact of an intervention on an outcome that will happen in the future. However, observational studies may be either prospective or retrospective.

Controls

The use of a comparison group, or **control group**, is essential when designing a study and interpreting any research findings. For example, when assessing the causal role of a particular factor for a disease, the risk of disease should be considered both in those who are exposed and in those who are unexposed to the factor of interest (Chapters 15 and 16). See also 'Treatment comparisons' in Chapter 14.

Bias

When there is a systematic difference between the results from a study and the true state of affairs, bias is said to have occurred. Types of bias include:

- **Observer bias**—one observer consistently under- or over-reports a particular variable;
- **Confounding bias**—where a spurious association arises due to a failure to adjust fully for factors related to both the risk factor and outcome (see Chapter 34);
- **Selection bias**—patients selected for inclusion into a study are not representative of the population to which the results will be applied;
- **Information bias**—measurements are incorrectly recorded in a systematic manner; and
- **Publication bias**—a tendency to publish only those papers that report positive or topical results.

Other biases may, for example, be due to **recall** (Chapter 16), **healthy entrant effect** (Chapter 15), **assessment** (Chapter 14) and **allocation** (Chapter 14).

[1] Hill, AB. (1965) The environment and disease: association or causation? *Proceedings of the Royal Society of Medicine*, **58**, 295.

13 Study design II

Variation

Variation in data may be caused by known factors, measurement 'errors', or may be **unexplainable random variation**. We measure the impact of variation in the data on the estimation of a population parameter by using the standard error (Chapter 10). When the measurement of a variable is subject to considerable variation, estimates relating to that variable will be imprecise, with large standard errors. Clearly, it is desirable to reduce the impact of variation as far as possible, and thereby increase the precision of our estimates. There are various ways in which we can do this.

Replication

Our estimates are more precise if we take replicates (e.g. two or three measurements of a given variable for every individual on each occasion). However, as replicate measurements are not independent, we must take care when analysing these data. A simple approach is to use the mean of each set of replicates in the analysis in place of the original measurements. Alternatively, we can use methods that specifically deal with replicated measurements (see Chapters 41 and 42).

Sample size

The choice of an appropriate size for a study is a crucial aspect of study design. With an increased sample size, the standard error of an estimate will be reduced, leading to increased precision and study power (Chapter 18). Sample size calculations (Chapter 36) should be carried out before starting the study.

Particular study designs

Modifications of simple study designs can lead to more precise estimates. Essentially we are comparing the effect of one or more '**treatments**' on **experimental units**. The experimental unit is the smallest group of 'individuals' which can be regarded as independent for the purposes of analysis, for example, an individual patient, volume of blood or skin patch. If experimental units are assigned randomly (i.e. by chance) to treatments (Chapter 14) and there are no other refinements to the design, then we have a **complete randomized design**. Although this design is straightforward to analyse, it is inefficient if there is substantial variation between the experimental units. In this situation, we can incorporate **blocking** and/or use a **cross-over design** to reduce the impact of this variation.

Blocking

It is often possible to group experimental units that share similar characteristics into a homogeneous **block** or **stratum** (e.g. the blocks may represent different age groups). The variation between units in a block is less than that between units in different blocks. The individuals within each block are randomly assigned to treatments; we compare treatments within each block rather than making an overall comparison between the individuals in different blocks. We can therefore assess the effects of treatment more precisely than if there was no blocking.

Parallel versus cross-over designs (Fig. 13.1)

Generally, we make comparisons between individuals in different groups. For example, most clinical trials (Chapter 14) are **parallel** trials, in which each patient receives one of the two (or occasionally more) treatments that are being compared, i.e. they result in *between-individual* comparisons.

Because there is usually less variation in a measurement within an individual than between different individuals (Chapter 6), in some situations it may be preferable to consider using each individual as his/her own control. These *within-individual* comparisons provide more precise comparisons than those from between-individual designs, and fewer individuals are required for the study to achieve the same level of precision. In a clinical trial setting, the **cross-over** design[1] is an example of a within-individual comparison; if there are two treatments, every individual gets each treatment, one after the other in a random order to eliminate any effect of calendar time. The treatment periods are separated by a **washout period**, which allows any residual effects (**carry-over**) of the previous treatment to dissipate. We analyse the difference in the responses on the two treatments for each individual. This design can only be used when the treatment temporarily alleviates symptoms rather than provides a cure, and the response time is not prolonged.

Factorial experiments

When we are interested in more than one factor, separate studies that assess the effect of varying one factor at a time may be inefficient and costly. **Factorial designs** allow the simultaneous analysis of any number of **factors** of interest. The simplest design, a 2×2 factorial experiment, considers two factors (for example, two different treatments), each at two **levels** (e.g. either active or inactive treatment). As an example, consider the US Physicians Health study[2], designed to assess the importance of aspirin and beta carotene in preventing heart disease and cancer. A 2×2 factorial design was used with the two factors being the different compounds and the two levels of each indicating whether the physician received the active compound or its placebo (see Chapter 14). Table 13.1 shows the possible treatment combinations.

We assess the effect of the level of beta carotene by comparing patients in the left-hand column to those in the right-hand column. Similarly, we assess the effect of the level of aspirin by comparing patients in the top row with those in the bottom row. In addition, we can test whether the two factors are **interactive**, i.e. when the effect of the level of beta carotene is different for the two levels of aspirin.

Table 13.1 Active treatment combinations.

	Beta carotene	
Aspirin	No	Yes
No	None	Beta carotene
Yes	Aspirin	Aspirin + beta carotene

[1] Senn, S. (1993) *Cross-over Trials in Clinical Research*. Wiley, Chichester.
[2] Steering Committee of the Physician's Health Study Research Group. (1989) Final report of the aspirin component of the on-going Physicians Health Study. *New England Journal of Medicine*, **321**, 129–135.

If the effects differ, we then say that there is an **interaction** between the two factors (Chapter 34). In this example, an interaction would suggest that the combination of aspirin and beta carotene together is more (or less) effective than would be expected by simply adding the separate effects of each drug. This design, therefore, provides additional information to two separate studies and is a more efficient use of resources, requiring a smaller sample size to obtain estimates with a given degree of precision.

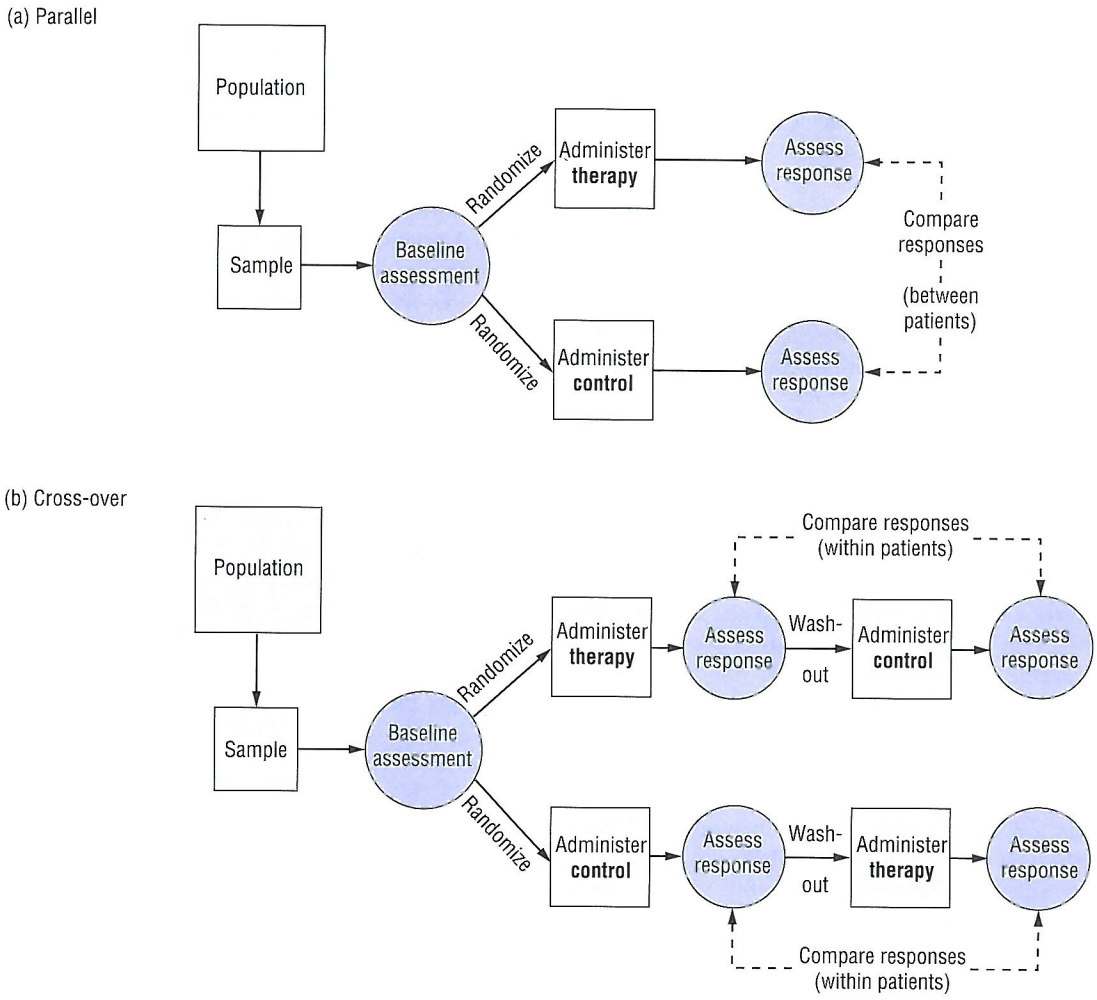

Figure 13.1 (a) Parallel, and (b) Cross-over designs.