RESEARCH METHODS AND MEASUREMENT

2

TWO

RESEARCH METHODS AND MEASUREMENT
and: **study design**
cross-sectional
longitudinal
experimental
qualitative
psychometrics

Good study design is essential if we are to be confident in our data and in the conclusions that we draw from them, and accurate measurement is an integral part of this. In this chapter we describe the importance of study design, and outline the various primary study designs which are used in health psychology and more widely, including cross-sectional, longitudinal and experimental designs. Each of these has its own strengths and weakness, and is ideally suited to specific kinds of research question. In addition, qualitiative methods are discussed, which offer the potential to answer a completely different kind of research question to quantitative methods, and serve a complementary function. Many important research questions can only be answered meaningfully using qualitative techniques, and good research will know which study design is best suited to answering a specific question. Finally, we discuss the science of psychometrics, which is concerned with ensuring that measurements are accurate and do indeed measure what they are intended to measure.

Key Concepts in Health Psychology

RESEARCH METHODS AND MEASUREMENT
and: STUDY DESIGN

MEANING   The **design** of the study refers to the means by which the research question will be addressed, specifically in relation to the data that will be collected, the comparisons that will be made, the experimental conditions (if any) that will be manipulated, and so on. In many ways, the design of a study is more important than the analysis of the data resulting from that study; if data are poorly analysed it will always be possible to re-analyse them, but if a study if poorly designed in the first place then it may never be possible to meaningfully interpret the data which result from it. The design of a study will also have an impact on how data are subsequently analysed.

ORIGINS   Research can be defined as belonging to one of two primary categories: **observational** studies and **experimental** studies. As suggested by the name, in observational studies there is no direct manipulation of variables within the study, and data are simply collected on groups of participants. In an observational study, while the research collects information on the attributes or outcomes of interest, no effort is made to influence these. An example might be the prevalence of a particular health behaviour (e.g. cigarette smoking) in different groups defined by socioeconomic status. In an experimental study, on the other hand, the experimenter directly influences events, in order to draw conclusions regarding the impact of that manipulation on the resulting observations. An example might be the impact of an intervention (e.g. an increase in self-efficacy beliefs) compared to a control condition on a particular health behaviour (e.g. giving up smoking).

A further sub-division which may apply to observational studies is between **prospective** studies, where information is collected about events subsequent to recruitment into the study, and **retrospective** studies, where information is collected about events in the past. Data regarding past events may be obtained from more objective sources (e.g. school records or patient notes) or, arguably, from more subjective sources (e.g. by self-report by the participant). Note that, in nearly all cases, experimental studies are prospective (even if the period of prospective study may be very short!), whereas observational studies may be retrospective or prospective. It should also be noted that whereas different treatments can be assessed retrospectively, this would *not* count as an experimental study, since the delivery of different treatments would not be an element of a pre-specified study. In this case, the study would be a retrospective observational study.

Another sub-division is that studies may be classified as **longitudinal** studies or **cross-sectional** studies. In longitudinal studies, changes over time are investigated. Again, note that, in nearly all cases, experimental studies are longitudinal (although once again, the period of longitudinal study may be very short), whereas
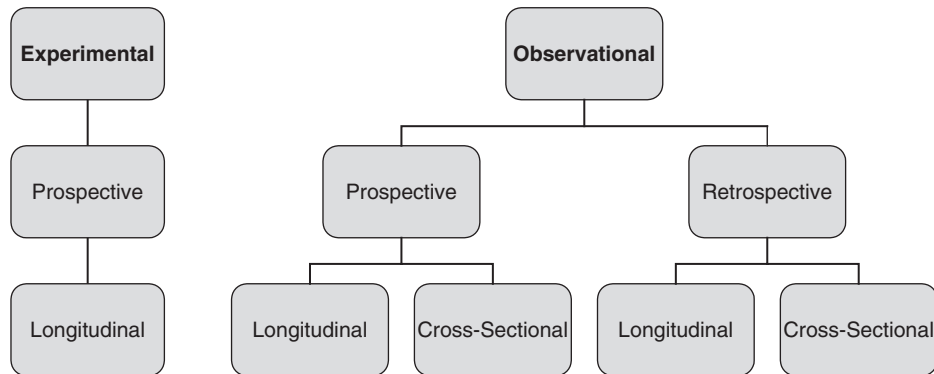
20

Figure 2.1   Study design

What are the basic categories of research design? The classifications that can be used include Observational versus Experimental, Prospective versus Retrospective, and Longitudinal versus Cross-Sectional. The distinction between observational and experimental relates to the purpose of the study, and is the most important distinction. The other two distinctions relate to the way in which the data are collected. Most of the combinations of these terms are possible, although not all are. These are the basic types of research design. Some of these categories are described in more detail in subsequent sections.

observational studies may be longitudinal or cross-sectional. In a cross-sectional study participants are observed only once, offering a 'snapshot' of the characteristics of interest at that particular moment. Most surveys, for example, are cross-sectional, although a variant is the **pseudo-longitudinal** design, where data on participants are collected at only one point in time, but this is done, for example, on, participants of different ages in order to indirectly construct a distribution of changes in a population with age. This approach is usually less resource-intensive than a proper longitudinal design, although somewhat more prone to giving erroneous results.

Finally, some kind of comparison is desirable in observational studies, and essential in experimental studies. This comparison group or condition is called the **control** group, to which the experimental procedure is not applied in the case of an experimental study design, or against which the group of interest is compared in an observational study design. For example, if we want to evaluate whether or not a new treatment or intervention is effective, we would usually do this by means of a comparison of the group given the intervention against one *not* given the intervention (namely, a control group). The presence of a control group, in both experimental and observational studies, strengthens the inferences that can be made from the results of the study.

While more detailed distinctions can be made, these represent the main ways in which a study can be designed (see Figure 2.1).

Key Concepts in Health Psychology

CURRENT USAGE The essence of statistical analysis is the understanding of likely causes of variation, which in the case of health psychology usually corresponds to variation in behavioural or psychological characteristics. Study design is important because it is the means by which we can be more (or less) confident that the apparent causes of variation in our study are in fact the genuine causes of variation.

For example, if we find that an intervention designed to improve self-efficacy beliefs is effective, we would need to know that those given the intervention demonstrated higher self-efficacy beliefs than those not given the intervention (namely, the control group), and that these two groups did not differ in their self-efficacy beliefs *before* the intervention was given. The strength of experimental study designs is that as long as participants are randomly allocated to either the experimental or control group, so that these groups do not differ at the outset of the study, any differences by the end of the study can be argued, with reasonable confidence, to be due to the experimental intervention. We would hope that there were no chance differences between the two groups *before* they received the intervention that might explain the difference observed after the intervention had been given.

Observational study designs, on the other hand, are more problematic as they require a comparison to be made where there is no active manipulation of events. For example, we may wish to investigate factors which it is impossible (or unethical) for us to manipulate directly (such as genetic variation). In this case, we would want to compare those with a particular characteristic (e.g. a particular genotype) to those with a different characteristic (e.g. a different genotype) on some measurement of interest (e.g. risk of a certain disease). Again, the more confident that we can be that our two groups do not differ in any other way, the more confident we can be that the difference of interest accounts for any observed differences in the measurement we take (in this example that having a particular genotype influences the risk of having a certain disease).

SIGNIFICANCE TO HEALTH PSYCHOLOGY Health psychology is the application of psychological and behavioural theory and research to the understanding of health, illness and healthcare. In many cases the factors influencing health, illness and so on are not amenable to experimental manipulation, so that observational study designs are frequently necessary. It is therefore critically important to understand (a) how best to design an observational study in order to maximize the confidence that one may have in the conclusions drawn from the results of the study, and (b) what limitations may exist in the inferences that may be drawn.

For example, we might be interested in whether or not personality influences alcohol consumption. A cross-sectional observational study might indeed reveal that people with high levels of trait anxiety do indeed drink more alcohol. However, we would not be able to say from these data alone whether high levels of trait anxiety eventually lead to higher alcohol consumption, or high alcohol consumption eventually leads to higher

levels of trait anxiety. The direction of **causation** would be impossible to infer from these results alone (although we might be able to speculate as to which alternative explanation was most plausible). Indeed, it might be the case that *both* effects are operating to give rise to our data (as a kind of positive feedback loop), but we would have no way of knowing.

Further reading

Altman, D.G. (1991) *Practical Statistics for Medical Research.* London: Chapman & Hall.

Chapter 5 discusses study design issues, while the rest of the book is an excellent introduction to medical statistics (which may be valuable to health psychologists who have covered statistics from the perspective of psychology only in the past).

Concato, J. (2004) Observational versus experimental studies: what's the evidence for a hierarchy? *NeuroRx,* 1*,* 341–347.

An interesting review which questions the widely-held assumption that experimental study designs are necessarily superior to observational study designs. Focused on bio-medical research, but applicable much more widely.

Sterne, J.A. and Davey Smith, G. (2001) Sifting the evidence – what's wrong with significance tests? *British Medical Journal,* 322*,* 226–231.

Very informative review of the dangers of emphasizing *p*-values in research papers, and in particular the focus on so-called 'positive' findings. Also discusses the role of study design in the validity and veracity of research findings.

See also **research methods and measurement and cross-sectional; research methods and measurement and longitudinal; research methods and measurement and experimental**

RESEARCH METHODS AND MEASUREMENT
and: CROSS-SECTIONAL

One of the most common and well-known study designs is the cross-sectional study design. In a cross-sectional study participants are observed only once, offering a 'snapshot' of the characteristics of interest at that particular moment. In this study **MEANING**

design, a population, or a sample thereof, is studied in a single instance, either by examining a single well-defined group of individuals, or by comparing cases (namely, those with an identifiable clinical condition) versus controls (namely, healthy individuals).

ORIGINS   Cross-sectional study designs are widely employed in epidemiological research, where characteristics of individuals (e.g. those presenting with a particular disease) are related to potential risk factors, or characteristics of cases (e.g. those with a particular disease) compared to the characteristics of controls (usually healthy individuals drawn from the general population). An example might be that cholesterol levels are higher among patients with coronary heart disease compared to healthy controls. Cross-sectional studies tend to be faster and cheaper to conduct than **longitudinal** studies, for the obvious reason that participants need only be studied at a single time point, rather than repeatedly over several time points.

CURRENT USAGE   It may be tempting to interpret the results of a cross-sectional study as if they were drawn from a **longitudinal** study. For example, if we look at cholesterol levels and the proportion of dietary calories derived from saturated fat in a cross-sectional study of healthy individuals, we might expect to find a very strong correlation. In this case, one might be tempted to conclude that high levels of saturated fat in one's diet *cause* increased cholesterol levels. This would be an example of treating cross-sectional data as longitudinal data, because one cannot make inferences about **causation** from cross-sectional data, however tempting it might be to do so (and even if one turned out to be correct!). A trivial example illustrates this: height and weight are very highly correlated, but it doesn't make sense to suggest that height *causes* weight, or indeed vice versa. Certainly, you would not want to suggest (at least to adults) that they should eat more in order to grow taller.

In a **case-control** study, the characteristics of a set of cases (that is, those with some condition, such as a disease) are compared to the characteristics of a set of controls (namely, those without the condition). For example, we might look at the smoking habits of those with and without lung cancer. Since we start out with a predetermined number of cases, the rarity of the disease is no longer an issue. As with all cross-sectional studies, case-control studies do not allow inferences to be made regarding causation. Even in cases where the risk factor appears relatively fixed (e.g. the presence or absence of a particular genetic mutation related to disease), one must still be cautious, since this risk factor may be operating directly (namely, causing the disease via a relatively direct biological pathway) or indirectly (namely, via effects on behaviours which themselves are related to disease).

While case-control studies may be informative in principle, they may be limited by the extent to which cases and controls are matched on other variables which are

not of interest. For example, if we are interested in comparing individuals with coronary heart disease to healthy controls, how do we go about selecting those controls? Do we simply randomly select individuals from the general population (who may, for example, have undiagnosed coronary heart disease)? Or do we select individuals from the general population who are *known* to not have coronary heart disease? Very slight differences between cases and controls may give rise to spurious results. For example, the rate of coronary heart disease is higher in males than in females, so that if the proportions of males among cases and controls differ slightly, we might see a difference that is independently related to sex (e.g. height) which *appears* to be related to coronary heart disease, but in fact is not in any meaningful sense.

A very large proportion of studies within health psychology are cross-sectional in nature. In essence, many do not differ in any important way from more traditional epidemiological studies which attempt to identify possible risk factors for disease. Whereas epidemiological studies generally focus on environmental risk factors (e.g. exposure to cigarette smoke), health psychology studies tend to focus on psychological and behavioural risk factors (e.g. self-efficacy beliefs, coping styles and so on).

SIGNIFICANCE TO HEALTH PSYCHOLOGY

The rationale behind the different studies is the same and, as a result, the limitations in interpretability are the same. For example, one might find a relationship between disease status and a particular coping style in a cross-sectional study, but one could still not be certain whether the coping style caused the disease, or vice versa. To do this, one would need to use a longitudinal study design and follow individuals with different coping styles over time, to see which of them subsequently developed the disease (and whether the development of the disease was linked to a change in coping style).

Further reading

Altman, D.G. (1991) *Practical Statistics for Medical Research.* London: Chapman & Hall.
Chapter 5 discusses study design issues, while the rest of the book is an excellent introduction to medical statistics (which may be valuable to health psychologists who have covered statistics only from the perspective of psychology in the past).

Biesheuvel, C.J., Grobbee, D.E. and Moons, K.G. (2006) Distraction from randomization in diagnostic research. *Annals of Epidemiology, 16,* 540–544.
Argues that experimental studies are not always necessary (in the context of diagnostic tests for disease), and that cross-sectional studies may be sufficient in some cases.

See also **research methods and measurement and study design**

RESEARCH METHODS AND MEASUREMENT
and: **LONGITUDINAL**

MEANING Another broad class of study design is the longitudinal design. In a longitudinal study a group of participants (**cohort**) is observed more than once (in contrast to a **cross-sectional** study, where participants are observed only once), and progress over time is assessed (e.g. the development or onset of disease). This allows changes over time to be recorded, and these to be related to characteristics or factors that existed and were measured prior to the onset of a behaviour or disease, for example. This might give information about risk factors for the development of a disease in a cohort of participants, some of whom developed the disease over time and some of whom did not.

ORIGINS **Longitudinal** study designs are widely employed in epidemiological research, and are generally regarded as being preferable to **cross-sectional** study designs as they are in principle less prone to potential problems of confounding arising from the inadequate matching of cases and controls in cross-sectional studies. Predictor variables are measured before the outcomes of interest occur, although it is also necessary to collect data on a large number of other variables to control for potential confounding. For example, depression levels at baseline may predict outcomes of interest, such as future illness, but may also be higher in females than in males, so that in order to identify the specific contribution of depression it would be necessary to measure and control for participant sex.

CURRENT USAGE The most important advantage of longitudinal study designs over case-control designs is that they allow slightly more certainty that factors which predict the outcomes of interest (namely, factors measured before the onset of the outcome which are significantly associated with the likelihood of that outcome subsequently occurring) that are genuinely associated with them. They do not allow inferences to be made regarding **causation** – this requires an **experimental** study design – but they do give stronger grounds for believing that the association between risk factor and outcomes is genuine as compared to case-control studies.

Nevertheless, there are still potential problems in longitudinal study designs. Not least, it can be difficult to determine whether a baseline factor which is associated with a particular outcome is in fact operating via some other, unmeasured, variable. For example, depression at baseline may predict the risk of future illness, but this may operate relatively directly (e.g. via effects on immune function) or indirectly (e.g. via effects on health behaviours such as diet and alcohol consumption). Although careful and comprehensive measurement of these potential confounding variables can help to exclude these possibilities, it will always remain the case that the association may be operating via some third, unmeasured variable. This is

complicated further by the possibility that some unmeasured variable may be causing both variation in the baseline predictor and variation in the outcome with which it is associated (so that the baseline measure is not causally associated with outcome at all). For example, the same genetic factors may be linked to both depression and illness, giving the appearance of a relationship between depression and illness. The fact that the relevant illness outcomes have not yet developed at baseline may simply be because these take longer to be expressed.

At a practical level, these studies are time-consuming and expensive to conduct, in particular if the outcomes of interest take several years to develop and be expressed. This long term follow-up also gives rise to potential problems of **attrition** within the cohort – participants may drop out, move, die, or otherwise be unavailable at follow-up, so that the final cohort for analysis may be considerably smaller than the one originally recruited. If the likelihood of dropping out of the study is related to the outcomes of interest, this may lead to biased results. For example, in a study of smoking cessation, individuals who are not successful in giving up smoking may be less motivated to remain in the study, resulting in successful quitters being over-represented in the final sample who complete the study.

Longitudinal studies are also not always an effective way to study associations, particularly when an outcome such as a particular disease is rare or takes a long time to develop. In the case of certain outcomes, there will simply not be enough individuals within a cohort who will develop the outcome of interest (or they will take a very long time to develop it). This will add to the cost and complexity of a longitudinal study design, as it will require an extremely large study (in order for a sufficient number of individuals to develop the outcome of interest) and/or a study which is continued for a very long period of time (in order to allow sufficient time for the outcomes of interest to develop). In situations such as these it is much more cost-effective and practical to run a case-control study instead.

A large number of studies in health psychology use longitudinal designs, although fewer than use case-control designs (primarily for reasons of cost and convenience). As with cross-sectional studies, these tend to be similar in many ways to traditional epidemiological studies of risk factors, but also tend to focus on psychological and behavioural risk factors (e.g. self-efficacy beliefs, coping styles and so on).

SIGNIFICANCE TO HEALTH PSYCHOLOGY

Longitudinal study designs in health psychology have a particular advantage over case-control designs because the psychological constructs of interest to many health psychologists are likely to be particularly affected by the onset of disease. For example, illness may result in a change in self-efficacy beliefs or coping styles, so that case-control studies will be very poorly suited to determining whether a particular copying style, for example, leads to illness, or vice versa. Of course, longitudinal study designs will still be limited, for the reasons described above, but will

at least be able to show whether variation in coping style precedes and predicts the onset of illness, in this hypothetical example.

Further reading

Altman, D.G. (1991) *Practical Statistics for Medical Research.* London: Chapman & Hall.
Chapter 5 discusses study design issues, while the rest of the book is an excellent introduction to medical statistics (which may be valuable to health psychologists who have covered statistics only from the perspective of psychology in the past).

Manolio, T.A., Bailey-Wilson, J.E. and Collins, F.S. (2006) Genes, environment and the value of prospective cohort studies. *Nature Review Genetics*, 7, 812–820.
Highlights specific research questions (in the context of genetic risk factors for disease) which longitudinal study designs are better suited to answering than case-control designs (see also Molecular Genetics).

See also **research methods and measurement and study design**

RESEARCH METHODS AND MEASUREMENT
and: **EXPERIMENTAL**

MEANING Experimental study designs are the strongest of the available study designs in that they allow inferences regarding **causation** to be made. In these studies, participants are randomly assigned to one of two (or more) conditions. Since allocation is random, any differences between the groups after exposure to the conditions can be inferred to be a result of the effects of the conditions. An example is a **randomized clinical trial** (RCT) of a novel medication compared to a placebo – if outcome measures are improved in the medication group compared to the placebo group, we can infer that the benefit is the result of the medication.

ORIGINS Experimental study designs arose from two independent research traditions. In bio-medical research, the rapid increase in the number of potentially effective medications and interventions available resulted in a need to be able to show experimentally that these interventions were more effective than simply a placebo. This resulted in the adoption of the randomized placebo-controlled trial as the gold standard of evidence that an intervention is an improvement over background effects such as spontaneous recovery and the **placebo effect**. The use of **randomization**, if done correctly,

ensures that participants in the intervention and control groups are similar at the start of the trial.

At the same time, the development of psychology and other behavioural sciences which investigated human and animal behaviour, resulted in the use of experimental study designs in laboratory contexts. In principle, the rationale behind the adoption of experimental designs is the same as in biomedicine: by allocating participants randomly to one of two (or more) groups, any resulting differences in behaviour can be inferred to have been caused by the experimental manipulation. For example, the impact of information on healthy eating (compared to a control condition which lacks this 'active ingredient', such as information on an unrelated topic) can be investigated in relation to the subsequent response to advertisements for food in an experimental, laboratory, setting.

An experimental study design is considered to provide the best level of evidence, both in the development of theory and understanding of mechanisms, and in the development of new or modified clinical practice guidelines. The principal reason for this is that only an experimental study design can demonstrate causality. A further advantage of experimental study designs is that data from these studies are the simplest to analyse statistically. Generally, any experimental design includes the following stages: recruitment of a sample from the relevant population, initial baseline assessment, randomization to the test or control condition, and follow-up assessment (see Figure 2.2).

CURRENT USAGE

Since the validity of the results from an experimental study depends on the randomization procedure not biasing the allocation of participants to either the test or control conditions, the quality of the randomization procedures used is critical. Simple methods may include the use of a coin toss or sealed envelopes (within which the condition allocation is given), while more sophisticated methods include the use of computer-generated randomization sequences and treatment allocation. The most important factor in randomization is that the method is not open to abuse. For example, if the experimenter can manipulate the allocation (perhaps to ensure that a patient he or she perceives to be in most need of treatment is not allocated to the placebo group) then the randomization procedure is subject to **bias** and the results of the study are invalid.

A further procedure which helps to ensure the validity of experimental study design is the **blinding** of experimenters and participants. Blinded means blind with respect to condition allocation. In a **single blind** study, participants do not know what condition they have been allocated to. This ensures that their responses will not be affected by prior expectations and that subsequent behaviour will not be affected by knowledge of the condition. In some rare instances, single blind refers to situations where participants know their condition allocation and only the person evaluating them is blinded. In a **double blind** study, participants and anyone who has contact with them or makes judgments about them is blinded to the assignment of condition. This ensures subjective judgments will not be affected by
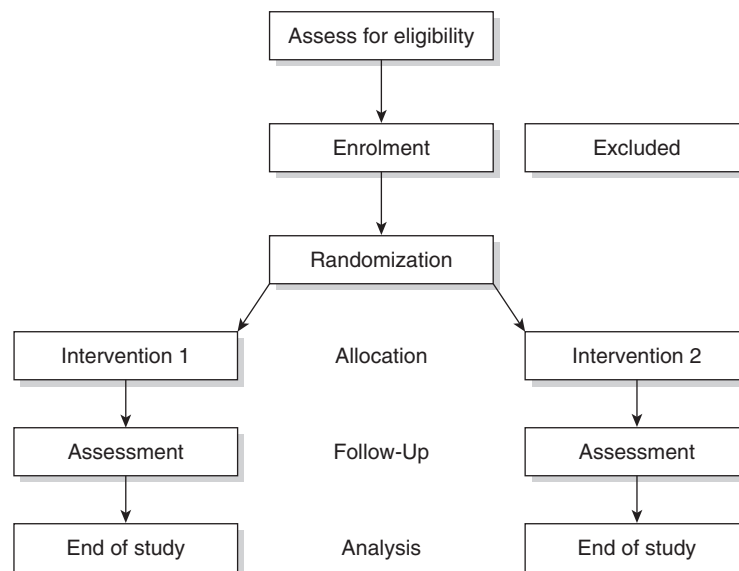
Figure 2.2   Experimental study design

Experimental study designs include initial assessment for eligibility, enrolment (at which points some participants may be excluded, for example because they decide that they no longer wish to take part) and randomization followed by allocation (on the basis of this randomization) to a particular experimental condition. After randomization there is a follow-up period where appropriate data are collected and analysed. Randomized controlled trials typically report the numbers of participants at each stage of this process in a diagram similar to that shown here.

knowledge of a participant's treatment. It is also possible to conduct **triple blind** studies where the individual responsible for data analysis is blinded to the relationship between the data for analysis and condition allocation within these data.

It is also important that the sample on which an experimental study is conducted is representative of the population about which one wishes to make inferences. In practice, this can be hard to achieve. For example, many laboratory studies recruit participants from the student population for convenience (because these studies are often conducted within a university's research environment), but the results of these studies may not generalize to the wider population. In RCTs, participants are often excluded if they have a pre-existing disease, or are on medication and so on. While this is done for reasons of safety when testing novel interventions, it once again means that the results may not generalize to the wider population on whom one might wish to implement the intervention in future.

A more complex experimental design is one which employs a **cross-over** design (see Figure 2.3), so that participants act as their own control. The basic elements of this variant are the same, but rather than being randomized to one condition

```
                    ┌─────────────────────┐
                    │ Assess for eligibility │
                    └─────────────────────┘
                              │
                              ▼
         ┌──────────────┐          ┌──────────────┐
         │  Enrolment   │          │   Excluded   │
         └──────────────┘          └──────────────┘
                │
                ▼
         ┌──────────────┐
         │ Randomization │
         └──────────────┘
           ╱            ╲
  ┌──────────────┐  Allocation  ┌──────────────┐
  │   Order 1    │              │   Order 2    │
  └──────────────┘              └──────────────┘
          │                            │
          ▼                            ▼
  ┌──────────────┐              ┌──────────────┐
  │ Intervention 1 │            │ Intervention 2 │
  └──────────────┘              └──────────────┘
          │                            │
          ▼                            ▼
  ┌──────────────┐  Follow-Up 1 ┌──────────────┐
  │  Assessment  │              │  Assessment  │
  └──────────────┘              └──────────────┘
          │                            │
          ▼                            ▼
  ┌──────────────┐   Washout    ┌──────────────┐
  │              │              │              │
  └──────────────┘              └──────────────┘
          │                            │
          ▼                            ▼
  ┌──────────────┐              ┌──────────────┐
  │ Intervention 2 │            │ Intervention 1 │
  └──────────────┘              └──────────────┘
          │                            │
          ▼                            ▼
  ┌──────────────┐  Follow-Up 2 ┌──────────────┐
  │  Assessment  │              │  Assessment  │
  └──────────────┘              └──────────────┘
          │                            │
          ▼                            ▼
  ┌──────────────┐   Analysis   ┌──────────────┐
  │ End of study │              │ End of study │
  └──────────────┘              └──────────────┘
```
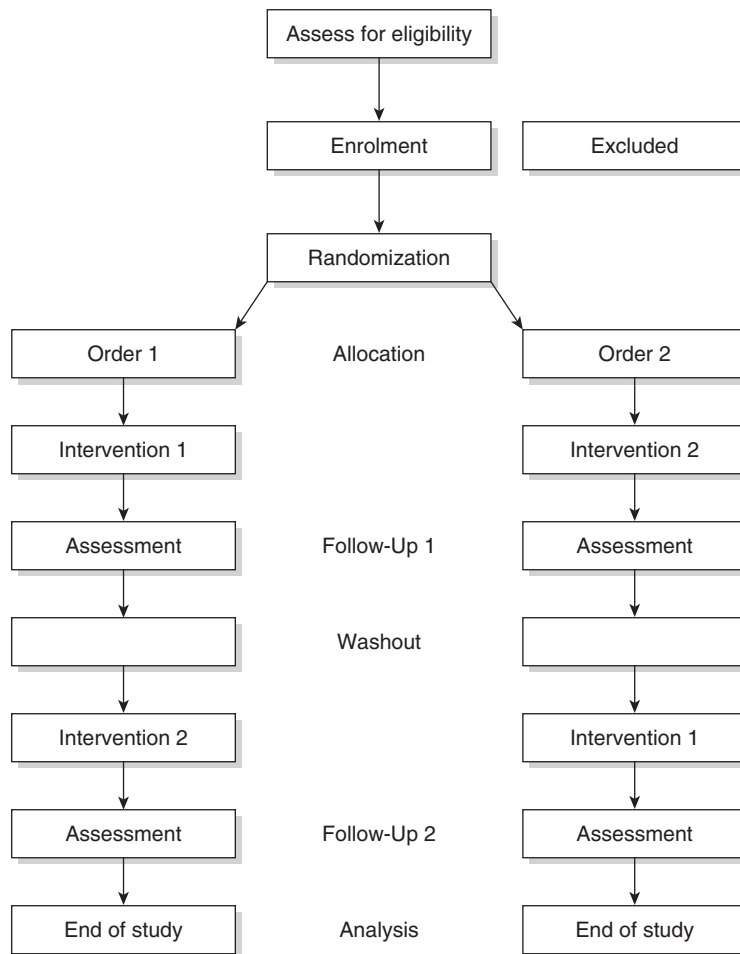
Figure 2.3   Cross-over study design

Cross-over study designs are the same as basic experimental study designs, except that each participant undergoes each intervention, usually with a washout period in between to allow the effects of the first intervention to dissipate. Participants are randomized to receive the two interventions in a particular order.

only, each participant receives both conditions, in sequential order. Participants are still randomized, but in this case according to the order in which they will receive the two conditions. Since participants are acting as their own controls this method reduces the impact of chance differences between participants in each condition. However, since there may be carry-over or practice effects (namely, completing follow-up questionnaires twice rather than once may influence results on the second occasion, perhaps due to participants remembering how they responded the first

time), the order in which conditions are administered to participants needs to be taken into account in the eventual statistical analysis. This means that this design is very complex in practice.

In the context of RCTs, there are, generally speaking, two types of intervention trials: **superiority** trials and **equivalence** trials. The purposes of the two types of trials are what their names suggest. Superiority trials are performed to show that one of the interventions is superior to the other. That is, they are carried out to establish a difference between the interventions. Equivalence trials are performed to show that the interventions are, according to some definition, the same or equivalent. The vast majority of RCTs are of the former kind, generally using some kind of appropriate placebo in order to show that a particular intervention is of some benefit. In health psychology, the choice of an appropriate placebo can be difficult, since the intervention is usually behavioural or psychological in nature, so that an analogue of a pharmaceutical placebo (namely, a sugar pill) is not readily available. Often some kind of minimally active intervention is used as a placebo in such instances (e.g. relaxation treatment).

SIGNIFICANCE TO HEALTH PSYCHOLOGY Experimental study designs contribute to both the development of theory within health psychology, and the establishment of evidence that health psychology interventions are efficacious (or not!). Laboratory studies on volunteers, for example, may provide the basis for confirming the role of specific psychological constructs in governing behaviour (e.g. the effects of mood on a desire for cigarettes), while RCTs may provide a basis for confirming that psychological interventions are effective (e.g. behavioural support for smoking cessation).

Further reading

Altman, D.G. (1991) *Practical Statistics for Medical Research.* London: Chapman & Hall.
Chapter 5 discusses study design issues, while the rest of the book is an excellent introduction to medical statistics (which may be valuable to health psychologists who have covered statistics only from the perspective of psychology in the past).

Jorgensen A.W., Hilden J. and Gotzsche, P.C. (2006) Cochrane reviews compared with industry supported meta-analyses and other meta-analyses of the same drugs: systematic review. *British Medical Journal,* 333, 782.
Interesting systematic which suggests that the results of meta-analyses of randomized controlled trials sponsored by the pharmaceutical industry differ from those without this sponsorship. Discusses the role of bias in study design and reporting.

See also **research methods and measurement and study design**

RESEARCH METHODS AND MEASUREMENT
and: QUALITATIVE

MEANING

Qualitative research attempts to provide a rich, descriptive account of the same psychological and behavioural phenomena, via analyses of **textual accounts** – either written or spoken – in order to produce detailed **narrative** reports. Quantitative research, by contrast, objectively measures psychological phenomena to enable statistical analysis, reducing phenomena to numerical values, and applying statistical operations to these to test hypotheses. While quantitative research attempts to be as objective as possible and to control experimental conditions, qualitative research accepts (even welcomes!) the subjective nature inherent in using a qualitative approach.

ORIGINS

Qualitative research has a long research tradition, which derives largely from a realist perspective, and focuses on **thematic methods**. That is, it attempts to identify themes within narratives, first-person accounts and so on which describe real-life phenomena without sacrificing the richness and detail of these phenomena. It uses qualitative methods to analyse qualitative data, although the exact nature of these data may be very different to those collected in a quantitative study.

If language use is the phenomenon of interest, then text is regarded as the object of research, with the real-world operating as the background to this. This kind of approach is used, for example, in **Discourse Analysis**. If real-world phenomena, such as behaviours, are the object of research, then text is regarded as the lens or medium through which the real-world (as the object of interest) can be understood and investigated. This kind of approach is used in **Grounded Theory** and **Focus Groups**.

CURRENT USAGE

Describing qualitative research as a single kind of research does not do justice to the many conceptual and theoretical approaches which fall within this very broad category. One common theme, however, is that behaviour is not necessarily amenable to scientific study in the same way that physical matter is, for example, and that attempts to do this, while potentially valuable in some respects, will necessarily give rise to an impoverished understanding. **Constructivism** regards behaviour as a response to situational demands, rather than as a fixed attribute of an individual. This would mean that personality, for instance, may be located *between* people, rather than *within* people (Hampson, 1988). This clearly has important implications with respect to **psychometric** approaches to the measurement of personality traits. According to this view, behaviour is negotiated between participants in a social exchange, and operates as a function of the situational and intrapersonal requirements. Qualitative methods are valuable because they attempt to understand the nature of this exchange, rather than simply reduce it to single numbers

Table 2.1   Differences between Quantitative and Qualitative Research

	Quantitative	*Qualitative*
	Requirement	
Question	Hypothesis	Interest
Method	Control and randomization	Curiosity and reflexivity
Data collection	Response	Viewpoint
Outcome	Dependent variable	Accounts
	Ideal	
Data	Numerical	Textual
Sample size	Large (power)	Small (saturation)
Context	Eliminated	Highlighted
Analysis	Rejection of null	Synthesis

There are various important differences between quantitative and qualitative research, both in terms of the requirements of the study and the ideal data, sample size and analytical methods. Most importantly, quantitative and qualitative methods address very different kinds of research question (see Table 2.3).

Table 2.2   Strengths and weaknesses of Qualitative Research

Strengths	*Weaknesses*
Low constraints of tradition or method	Poor internal reliability
Grounded hypotheses	Weak decisiveness
Contextualized hypotheses	Poor generalizability
Non-normative focus	Rarely integrated
Comprehensiveness	Seems easy
Detail	Seems harmless

There are a number of strengths and weaknesses to qualitative research, some of which are listed here. The most important strengths include the comprehensiveness of qualititative research, which embraces richness and depth. Unfortunately, it is relatively uncommon for qualitative research to be integrated with quantitative research, even though it has the potential to inform the design of quantitative studies.

(such as the score on a personality questionnaire). Quantitative research emphasizes replicability, reliability and validity, while qualitative research emphasizes contingency, situatedness and reflexivity (see Table 2.1).

Data collection methods in qualitative research are different from those in quantitative research in a number of important ways. Observation, either as a participant or non-participant, is critical (as opposed to measurement via instruments), with **interviews** (either of individuals or groups) and **documentary analysis** (of records, transcripts, and so on) forming the basis of data collection.

A variety of qualitative methods of data collection and data analysis exist, the most common of which include grounded theory and discourse analysis. **Grounded Theory** attempts to construct a theory based on qualitative data, without pre-existing hypotheses. The core feature of Grounded Theory is simultaneous data

Table 2.3   Different research questions and methods in Quantitative and Qualitative Research

Quantitative question	Quantitative method	Qualitative question	Qualitative method
'By how much does thrombolytic therapy improve survival in acute myocardial infarction?'	Meta-analysis of randomized controlled trials	'Why do some patients delay seeking help when they have acute central chest pain?'	Individual (semi-structured) interviews, with cross-validation between interviewers
What proportion of smokers will give up when so by advized to do their doctor?'	Prospective observational (cohort) study	'What sort of smoker responds to advice to quit, and how might we influence the remainder?'	Individual (semi-structured) interviews, with cross-validation between interviewers
'What features of acute febrile illnesspredict serious diseases such as bacterial meningitis?'	Retrospective (case-control) study	'What worries parents when their pre-school children are acutely ill, and why?'	Individual (semi-structured) interviews, or focus group discussions

Quantitative and qualitative methods each have their place in research, and are each best suited to answering specific kinds of research question. Some examples are given here, along with the specific methods or study designs which might be used to answer each question.

Table 2.4   Lay beliefs about hypertension (Source: Heurtin-Roberts, 1993)

High Blood	*High Pertension*
Physical disease of the heart and blood	Disease of the 'nerves' caused by stress, worry and personality
Blood too 'hot', 'rich' or 'thick' over extended periods	Believed to be volatile and episodic, rather than long lasting
Level of blood rose slowly and remained high	Levels of blood rise and fall rapidly at times of stress
Caused by diet, heredity and 'heat' in the environment	Treated by amelioration of stress and psychological symptoms

Heurtin-Roberts investigated lay beliefs about hypertension among African-American women in New Orleans. This research revealed a complex pattern of beliefs related in part to modern bio-medical conceptions of hypertension, but also to cultural beliefs and folk medicine. Health education campaigns which fail to take into account people's existing beliefs about disease may not succeed if the advice provided does not match these existing beliefs.

collection and analysis, with analytic categories developed from the data as data collection is ongoing. Comparisons of data and data, data and concept, and concept and concept allow these categories to be developed, and further theoretical sampling and data collection are used to confirm these categories. Discourse Analysis is concerned with the role of language in the construction of social reality, and emphasizes discourse as performance, and in particular the fluidity and variability of discourse. It asks what participants are doing with discourse (that is, what function it serves for them), and investigates the creation of meaning in language. Other methods exist, including phenomenological analysis (which explores how we make sense of the world), narrative psychology (which explores the role of narratives in understanding the world), conversation analysis (which explores spontaneously generated conversation), and cooperative enquiry (where the researcher is a participant in the research).

Qualitative methods can be argued to be of more importance in health psychology than in other areas of psychology, because the concepts which health psychologists deal with are intended to reflect the 'real world' in a very direct way, which has application in the understanding and treatment of physical health and illness. It is possible (indeed likely) that the concepts employed in biomedicine may not be understood by the general public, so that interventions, health education campaigns and so on may fail because they are not understood by those they target.

One well-known example illustrates this potential problem. Heurtin-Roberts and colleagues (1993) investigated folk beliefs regarding hypertension in African-American women in New Orleans. The study identified two distinct groups of individuals in this population, who could be characterized according to the differences in their beliefs regarding this disease. One group believed in the bio-medical definition of hypertension, while other group believed in two distinct diseases (see Table 2.4).

From a bio-medical perspective, the two distinct 'diseases' identified by the second group of individuals each shares some features of the bio-medical concept of hypertension. Clearly, health education campaigns based on the bio-medical concept of hypertension which do not take into account these important and strongly-held folk beliefs may be unsuccessful.

Further Reading

Heurtin-Roberts, S. (1993). 'High-pertension' – the uses of a chronic folk illness for personal adaptation. *Social Science Medicine,* 37, 285–294.
Investigation into lay beliefs around hypertension, and the extent to which chronic illness can be understood in a cultural context which emphasizes particular coping responses and behaviours.

Smith, J.A. (ed.) (2003) *Qualitative Psychology: A Practical Guide to Research Methods*. London: Sage
An excellent introduction to qualitative methods, specifically as applied to psychology. Discusses a range of techniques and conceptual models and frameworks, including **Grounded Theory**, **Discourse Analysis** and so on in detail.

See also **research methods and measurement and study design**

RESEARCH METHODS AND MEASUREMENT
and: **PSYCHOMETRICS**

MEANING

Psychometrics is concerned with the measurement of psychological **traits** such as knowledge, abilities, attitudes and personality traits, typically although not necessarily using questionnaire-based methods. The study of psychometrics relates primarily to the investigation of differences between individuals and between groups of individuals. The principles of psychometrics guide the development and use of instruments and procedures used in the measurement of psychological traits.

ORIGINS

A number of prominent figures in the early days of psychological science contributed to the development of the field of psychometrics. The most well-known, at least with respect to their contribution to psychometrics, are Francis Galton and Charles Spearman, both of whom developed and refined psychometric approaches to the measurement of intelligence and general aptitude or cognitive ability. Their contribution

also included the development of factor analysis (see Box 2.1), which continues to be used extensively in psychometrics and plays a central role in the development of questionnaire instruments.

Box 2.1   Factor analysis

Suppose a psychologist proposes a theory that there are two kinds of intelligence, 'verbal intelligence' and 'mathematical intelligence'. Behavioural traits such as these cannot be observed directly and have to be measured indirectly, for example using an intelligence test. Factor analysis is used to identify 'factors' that explain a variety of results on different tests. For example, in our hypothetical examples, factor analysis could be used to identify whether performance on intelligence tests was related to performance on two different kinds of question within these tests (for example, those relating to 'verbal intelligence' and those relating to 'mathematical intelligence'). If these factors exist, people who get a high score on one test of verbal ability will also get a high score on other tests of verbal ability. Factor analysis is the method by which these inter-relationships are identified.

Factor analysis has a long history in psychometrics, and has been used to identify personal traits (for example, what is the optimal number of traits which we need to describe a large proportion of the variability in human behaviour) and the structure of intelligence (for example, what are the different sub-types of intelligence which exist, and how do they relate to each other). It has the advantage of providing structure and simplicity to complex datasets, although it has the disadvantage of being a complex procedure, and is only as good as the data which are used in the factor analysis.

Psychometrics has been applied extensively to the measurement of personality, attitudes, mood, intelligence, cognitive ability, and so on. The inherent difficulty in measuring these constructs, which lack a tangible physical correlate, drives the use of psychometrics, which attempts to properly quantify and define these constructs. The earliest psychometric instruments were designed to measure intelligence, the best known of which is the Stanford-Binet IQ test developed originally by Alfred Binet, and psychometric instruments continue to be used widely within education and educational psychology. Another major area of study in psychometrics relates to personality, and a large number of personality measures and related models and theories exists.

There are three fundamental properties of any good test or instrument: **objectivity**, **reliability** and **validity**. The traditional and most important concepts in psycho- metrics are reliability and validity. Reliability refers to the extent to which a measurement is free from unsystematic error, and therefore measures some- thing accurately. Validity is the extent to which a measurement indexes the underlying psychological construct which it is intended to. To illustrate this, a clock which is five minutes fast is reliable (it is measuring time), but invalid (it is not measuring the right time). Ideally, an instrument should also be free from systematic **error** (e.g. systematic biases, such as a ruler which includes 11 millimetres in each centimetre). However, it is unsystematic error which is the main concern, since a test cannot be valid if it is not reliable. It is, however, pos- sible for a test to be reliable but not valid (as in the case of the fast clock, above).

In technical terms, reliability refers to the proportion of variance in the measurement which is due to unsystematic (that is, random) error. Unsystematic error may arise from a number of sources. For example, if people's responses to a questionnaire vary depending on their mood (when the questionnaire is not in fact intended to measure variation in mood but something conceptually unre- lated, such as intelligence), then this variation is uninformative with respect to the intended measurement. Since this uninformative variation will vary unpre- dictably (namely, unsystematically) between participants, this will serve to reduce the overall reliability of the instrument. Various potential threats to reli- ability exist, which may be more or less relevant depending on exactly what the instrument in question is intended to measure. For example, instruments which are intended to measure stable, underlying traits (such as personality and intelli- gence) should be relatively unaffected by factors such as time of day, current mood and so on. Conversely, instruments which are intended to measure state factors (such as mood) are likely to be sensitive to these factors (given that mood varies from time to time).

Test-retest reliability refers to the extent to which the measurement taken on an instrument at one time point correlates with that taken at another time point, although this should only be high in cases where we do not expect the measure to change over time (e.g. intelligence). Other tests of reliability include inter-rater reliability (which assesses the extent to which different experimenters give the same rating, in cases where interview rather than ques- tionnaire instruments are used) and alternate form reliability (which assesses the extent to which different versions of a test are correlated, in cases where one might want to use two versions of a test to avoid practice or learning effects).

One sense in which all instruments should be relatively free from unsystematic error is that they should be internally consistent. That is, all items on a questionnaire should tap into the same underlying construct. This is typically assessed using the

split-half test, which calculates the correlation between one half of the items on a test with the other half. If all the items are measuring the same underlying construct, the two halves should be highly correlated. Cronbach's alpha is the mean of all possible split-half correlations, which is a more accurate measure of internal consistency.

The validity of psychometric instruments is difficult to assess in practice because the constructs so measured do not have tangible physical properties (unlike, say, blood pressure). The validity of such instruments, therefore, must be assessed indirectly, by correlating measures with a criterion measure known to be valid. Various forms of validity exist (unlike reliability, where there is only one kind, although *threats* to reliability may occur from different sources).

Concurrent validity is a sub-set of criterion validity, and refers to the extent to which the instrument correlates with an appropriate criterion which is measured at the same time (e.g. a new version of an intelligence test and an existing version). Predictive validity is another sub-set of criterion validity, and refers to the correlation with an appropriate criterion measured at a later time (e.g. intelligence and future examination performance). Content validity reflects the extent to which the instrument includes measures of all of the relevant features of the construct of interest (usually assessed by an appropriate panel of experts). Face validity reflects the extent to which the items of a test appear relevant and appropriate to (usually non-expert) participants who are likely to form the 'target audience' for the instrument. A measure has construct validity if it is related to other variables as required by the theory which underpins the instrument (e.g. intelligence and examination performance), and also unrelated to variables to which it should not be related (e.g. intelligence and personality). The former is sometimes known as convergent validity, and the latter as divergent validity, and the pattern of evidence from both provides evidence of construct validity. This is sometimes assessed using the Multi-Method Multi-Trait Approach.

SIGNIFICANCE TO HEALTH PSYCHOLOGY

Health psychology relies heavily on psychometric instruments, including mainly questionnaire instruments but also interview instruments, to quantify the constructs which are central to many of the research questions studied. It is therefore important that these measures are both reliable and valid – we would want to be reasonably certain, for example, that a measure of health-related locus of control was in fact measuring what it purported to measure. For this reason, researchers generally use instruments which have undergone an extensive and stringent process of reliability and validity testing. The use of instruments which have not gone through this process (e.g. questionnaires developed quickly on the basis of a researcher's intuition and solely for the purposes of a single study) is discouraged.
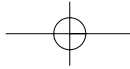
Further reading
Kline, P. (1999). *A Psychometric Primer.* London: Free Association Books.
Gives an introduction to test theory and the principles of psychometrics, with a particular emphasis on the questionnaire measures typically used by psychologists.

See also **research methods and measurement and study design**