QUESTIONNAIRE DESIGN, INTERVIEWING AND ATTITUDE MEASUREMENT

New Edition

A.N. Oppenheim

To
Betwyn

© 1966; 1992

Pinter Publishers
London and New York

It is true, of course, that there are occasions when we wish to study a finite, special group, for example all the women in Dr Z's practice who gave birth to a baby in year Y, or all the employees who left firm X in a given month. But even in these circumstances research design problems arise. Quite probably, we wish to study these particular mothers, or job leavers, because we want to generalize about them; or compare them to other such groups; or predict what may happen this year, or next; or use their responses in order to improve our procedures. But unless our study is designed appropriately and is aimed at the correct target population, we will be unable to draw any such conclusions or comparisons. There may also be non-response problems (see Chapter 7) so that some of the necessary information will be missing, and the remainder of the responses may be biased. These matters will present us with further research design problems.

The need for good design becomes even more acute when we wish to undertake a more ambitious study. Suppose, for example, that we are asked to evaluate some social intervention such as the use of road paints to warn motorists to slow down before an intersection or the institution of a telephone helpline for children who are being abused. We might be asked to say, on the basis of survey research, whether such interventions are achieving their goal/ are cost-effective/should be discontinued or changed and so on. How would we go about this? How would we try to make sure that our conclusions are valid and could form a sound basis for generalization and further action?

It might be helpful at this point to make a rough-and-ready distinction between research *design* and research *techniques*, although each influences the other to some extent.

The term research *design* here refers to the basic plan or strategy of the research, and the logic behind it, which will make it possible and valid to draw more general conclusions from it. Thus, the research design should tell us how our sample will be drawn, what sub-groups it must contain, what comparisons will be made, whether or not we shall need control groups, what variables will need to be measured (when and at what intervals), and how these measures will be related to external events, for example to social, medical or other interventions. Research design is concerned with making our problem researchable by setting up our study in a way that will produce specific answers to specific questions. Good research design should above all make it possible for us to draw valid inferences from our data in terms of generalization, association and causality.

Research *techniques*, on the other hand, are the methods used for data generation and collection. Shall we gather our data by interview, by telephone or by postal questionnaire? How shall we measure attitudes, purchasing behaviour, social integration, conservatism or friendship patterns? Can we put questions together in groups to form inventories and scales (see Chapter 9)? How shall we analyse the contents of the replies to our questions? How shall we deal with missing data? Essentially, research techniques are concerned with measurement, quantification and instrument building and with making sure that our instruments are appropriate, valid and reliable.

This distinction between research design and research techniques holds true for all scientific disciplines. Moreover, the principles of research design are applicable quite generally to experimentation, the requirements of sampling and the logic and standards of inference in any discipline. However the creation and application of measuring instruments and data collection techniques tend to be specific to each discipline or group of disciplines. Cross-disciplinary research, such as doctors conducting social surveys of their patients' smoking habits, requires familiarity with measuring techniques in more than one discipline.

Here we shall be concerned primarily with the *design* of social surveys, the planned architecture of inquiry. In later chapters we shall deal with measurement and instrument building, that is with research *techniques* such as scales and questionnaires. However, as we have indicated before, these two aspects of research are often interlinked. The design of the research will determine whom we should question and what questions we should ask, while technical and fieldwork problems may set constraints on the research design.

S tart

First steps in survey design

Too often, surveys are carried out on the basis of insufficient design and planning or on the basis of no design at all. 'Fact-gathering' can be an exciting and tempting activity to which a questionnaire opens a quick and seemingly easy avenue; the weaknesses in the design are frequently not recognized until the results have to be interpreted — if then! Survey literature abounds with portentous conclusions based on faulty inferences from insufficient evidence misguidedly collected and wrongly assembled. Not everyone realizes that the design of a survey, besides requiring a certain amount of technical knowledge, is a prolonged and arduous intellectual exercise in the course of which we are continuously trying to clear our own minds about our goals. We often find that, as the research takes shape, our aim undergoes a number of subtle changes as a consequence of greater clarity in our thinking (see Chapter 2). Such changes may require a new and better design, which in turn will lead to a better specification for the instruments of measurement.

The drawing up of the research design takes place at the very beginning of the research process, though the plan may have to be changed later. A social research study may last from a few months to many years, but most surveys go through the same stages or cycles of stages.

We may distinguish the following:

1. Deciding the *aims* of the study and, possibly, the theories to be investigated. General aims must then lead to a statement of specific aims, and these should be turned into *operationalized* aims; that is, a specified set of practical issues or hypotheses to be investigated. This should lead directly to a statement of the *variables* to be measured, and for each of these a set of questions, scales and indicators will have to be formulated.
2. Reviewing the relevant *literature*; discussions with informants and interested organizations.
3. Preliminary *conceptualization* of the study, followed by a series of exploratory or 'depth' interviews; revised conceptualization and research objectives (see below).
4. Deciding the *design* of the study and assessing its feasibility within the

limitations of time, costs and staffing. Abandon the study at this point, or reduce its aims, if the means to carry it out are insufficient.

5. Deciding which *hypotheses* will be investigated. Making these hypotheses specific to the situation (that is making the hypotheses operational). Listing the variables which will have to be measured. For instance, if we have some hypotheses about political participation, then how shall we operationalize this behaviour, and what variables shall we need to measure: party membership, fund-raising activities, going to political rallies, watching political events on TV, displaying a window poster or bumper sticker?

6. Designing, or adapting, the necessary *research instruments* and techniques such as postal questionnaires, interview schedules, attitude scales, projective methods, check lists or rating scales.

7. Doing the necessary *pilot work* (see Chapter 4) to try out the instruments, making revisions where necessary and trying them out again. Piloting other aspects of the research such as how to gain access to respondents.

8. Designing the *sample(s)*. Will the sample need to be representative (that is, a probability sample, see Chapter 3) and, if so, of whom? Are there lists or other sampling frames from which to draw a sample? Shall we need a control group? Shall we need a follow-up sample? How shall we cope with non-response and missing data, that is with differences between the designed sample and the achieved sample? Pilot work on the techniques of drawing the sample.

9. Drawing the sample: *selection of the people* to be approached.

10. Doing the *field-work*. This will involve the actual data collection process through interviewing or the sending out of the postal questionnaires; the day-to-day control of these complex operations; and the collection and checking of the returns. Contrary to expectations, the field-work stage is often shorter than the preliminary stages.

11. *Processing the data* (see Chapter 14): coding the responses, preparing the data for analysis and entering them into a computer if available.

12. Doing the *statistical analysis* (see Chapter 15), simple at first but becoming more complex; testing for statistical significance.

13. Assembling the results and *testing the hypotheses*.

14. Writing the *research report*: describing the results in words and tabulations; relating the findings to previous research; drawing conclusions and interpretations.

It should be clear from the above that a social survey is a complex operation, and a first requirement is therefore the development of a clear plan or over-all research design. It is the research design which must hold all the parts and phases of the enquiry together. The design must aim at precision, logic-tightness and efficient use of resources. A poorly designed survey will fail to provide accurate answers to the questions under investigation; it will leave too many loopholes in the conclusions; it will permit little generalization; and it will produce much irrelevant information, thereby wasting case material and resources.

Much will depend on the quality of the research conceptualization (see (3) above). 'Conceptualization' here refers to an improved and more detailed

statement of the study's objectives, preferably with theoretical underpinnings. If, for example, we intend to study the buying of home computers, should we question both owners and non-owners? What about ex-owners? And what should we ask them about: their income, the availability of computers at their place of work, various aspects of home computer use (domestic accounts, word processing, computer games), the way people make decisions about major domestic expenditures? Later, a re-conceptualization might add to this list, for example educational use (if they have children); 'keeping up with the Joneses' (prestige factors); typing skills; fears about new technology (which might be gender-related); leisure pursuits; health-related fears; and so on.

Here is an example of a poor design. The research took place in a major conurbation in a developing country which was attracting large numbers of migrants from distant rural areas. The researchers found that many of the more recent arrivals in the city were suffering from psychosomatic and stress-related disorders. They concluded that this was due to the new migrants being adversely affected by the problems of urban living. However, the researchers did not present any data about comparable samples of long-established city dwellers to show whether they were experiencing higher, lower or similar levels of stress-related disorders compared to the new migrants. Moreover, assuming the findings to be valid, they do not constitute proof that the stresses arose from city life. It is equally possible to suggest that these migrants were suffering from long-standing stress-related disorders which had been present before they came to the city. It may also have been the case that the migrants were motivated in part to move to the city because they found the difficult living conditions in the rural areas too stressful! Thus, in failing to incorporate a control or comparison group, the design left loopholes for at least two alternative interpretations of the findings: a pre-existing condition and selective migration. Also, the researchers were too uncritical in assuming certain disorders to be necessarily due to stress and urban living conditions. We should always remember that association is not proof of causation.

One way of testing the adequacy of a research design in advance of the field-work is to run through the natural sequence of survey stages *in reverse order*. For example, let us assume that in our final report (stage 14, the Research Report) we expect to be able to show whether men differ from women in the ways in which they are politically active. To demonstrate this we shall need some statistical tables and cross-tabulations, for example a check-list of political activities reported by our respondents, cross-tabulated by gender (the statistical analysis stage). At this point we must draw up some dummy tables, showing the relevant variables in cross-tabulation with certain sub-groups, for example gender. In order to generate such tabulations we must have asked questions about political activism and we must have administered such a check-list; we must also know the sex of each of our respondents (the research techniques and pilot work stages). To be able to generalize from our findings, we must have established an accurate relationship between the members of our sample and the target population (the sampling and sample design stages). And to be meaningful, the limited topic of gender differences must be part of a wider conceptualization or theoretical framework (the literature review and conceptualization stages).

Thus, by running backwards through the survey stages we can try to ensure logic-tightness, so that nothing is left out and we shall have measured everything that needs to be measured.

The formulation of any questionnaire to be used in a survey must be an integral part of the research design stage. A questionnaire is not just a list of questions or a form to be filled in. It is essentially a measurement tool, an instrument for the collection of particular kinds of data. Like all such instruments, the aims and specifications of a questionnaire stem directly from the overall research design. Such objectives and specifications are not always obvious from a perusal of the questions. However, we cannot judge a questionnaire unless we know what job it was meant to do. This means that we have to think not merely about the wording of particular questions but, first and foremost, about the design of the investigation as a whole.

Thus it is essential for any researcher to draw up a statement (a) of general aims for the proposed study and to develop these into (b) a set of specific aims; these should be further developed into (c) a set of operationalized aims; that is, specific issues or hypotheses, which should point directly to (d) a list of variables, scales or indicators which will be needed: for example, how shall we measure parental education or strictness of upbringing? Dummy tables for the proposed analysis should also be drawn up at this point.

An actual case

In the early fifties, when television-viewing was rapidly gaining ground in Great Britain, a survey was commissioned to supply answers to two broad questions: (1) what kinds of people were buying television sets? (2) How was the new medium affecting them? The large-scale survey that eventually took place used two samples. The first was intended to be representative of television-owners and was drawn from the lists of television licence holders. This sample was bound to be incomplete since it failed to include set owners who had not applied for a licence, but by and large it could give reasonable estimates of the income, family size, social and educational background, religion, consumer habits, and so on of law-abiding television owners at that time.

The second sample was assembled by asking the interviewers to question an adjacent non-television householder for each set owner in the first sample. It was reasoned that by interviewing the people next door a sample could be obtained, for purposes of comparison, that would be similar to the television owners in socioeconomic background, consumer habits and so forth.

The faults in this design were only realized toward the end of the investigation. Let us consider the logic of it step by step. The research had two purposes. The first was to find out what kinds of people were buying television sets. This implied something more than a purely descriptive study of a representative sample of set owners, since to state that such families had an average income of X pounds a week or that Y per cent of them had telephones or washing machines would be almost meaningless unless it could be put into some kind of context; in other words, it had to be compared to other figures. It might then become possible to conclude that television buyers more often came

from the lower income levels, or had larger families, were less religious and so on. For a few points of comparison, data were probably available for the nation as a whole (including the television owners), but most of the questions had to be asked afresh. In order to show how television buyers differed from non-buyers, the second sample would have to be a representative sample of all the people who had not bought a television set. That a sample of people living next door to set owners would accurately represent the non-buying public was clearly too much to hope for, since in any case they had been chosen on the grounds of similarity in environment and socioeconomic background to the set owners. As the figures subsequently showed, the second sample, not being representative of the non-television buyers, was useless for purposes of comparison. Consequently, the accurate figures obtained from the first sample could not be made to reveal any social change or pattern.

And what of the second purpose of the investigation; how was the new medium affecting set owners? Here there were many difficulties, since it is often hard to disentangle the effects of *one* variable, such as television, from the effects of many others operating at the same time. A comparative study can try to do this by choosing a control sample as similar as possible to the television sample in every relevant respect, so that we can see what the people in the television sample would have been like if they had not bought television sets. There are still weaknesses even then because of pre-existing differences, matching problems and effects of guest-viewing, but the sample design failed to produce sufficient similarity between the two groups. Consequently, there was no way of telling whether or in what way changes had been produced in the television sample.

The second sample, that of the neighbours, turned out to be useless. It was neither a representative sample of non-television buyers (as required for the first purpose) nor a closely comparable control sample (as required for the second purpose). It should have been realized from the beginning that, since the inquiry had two objectives, two *different* comparison samples were required. This may seem obvious in retrospect, but the professional survey organization that carried out the research realized too late what had happened. Mistakes like this occur every day and can only be prevented by making the greatest possible effort to clarify our objectives before we start.

Surveys versus experiments

How do we decide which of the many types of research design to choose for our particular purpose? This is a difficult question to answer since every research project presents its own particular design problems. Still, it is possible to discuss some basic types of research design in general terms and to make some rough-and-ready distinctions. In the next chapter we shall discuss cross-sectional designs, longitudinal designs, intervention designs, factorial designs and regression designs. All of these (and several others) can be grouped together under the heading 'analytic designs' (see Chapter 2) because, in one way or another, they attempt to deal with associations or with cause-and-effect relationships in ways analogous to lab experiments. They try to explain things, they seek to answer 'why' questions, or 'what determines X' questions and, in this way, they are to

be distinguished from the other broad category, the 'Descriptive' designs. (See Chapter 3.)

Beyond the world of survey research, the scientific endeavour is characterized by another type of analytic design, the controlled laboratory experiment. This method, which involves the systematic manipulation of variables, is not usually available to the survey researcher who cannot, say, make people want to adopt a pet to see what effect this will have on them or cause them to be burgled to see how this will influence their attitudes towards the police. While it is sometimes possible to conduct simple experiments with people involving the manipulation of several variables, it should not be assumed that the controlled laboratory experiment is the 'ideal' design. Experiments and surveys are often contrasted with each other. Those who favour surveys criticize experiments for being unrepresentative, for dealing with artificial situations and for often failing to achieve the degree of precision and control that might justify them. Experimentalists are critical of surveys because of their reduced ability to control or manipulate important variables, for following events rather than making them happen and for their inability to prove causal relationships. Nevertheless, the survey researcher using an analytic design will have taken many a leaf out of the experimentalist's book. It would be more helpful to suggest that choosing the best design or the best method *is a matter of appropriateness*. No single approach is always or necessarily superior; it all depends on what we need to find out and on the type of question to which we seek an answer. Indeed many research enquiries have employed surveys *and* experiments at different stages, using the result of the one to inform and refine the other, and so producing conclusions that are both precise and representative.

Descriptive versus analytic designs

As stated above, it is possible to make a broad distinction between two types of survey: (1) the descriptive, enumerative, census-type of survey; and (2) the analytic, relational type of survey.

The purpose of the descriptive survey is to count. When it cannot count everyone, it counts a representative sample and then makes inferences about the population as a whole. There are several ways of drawing a sample, and the problems of sampling, sampling bias and sampling error are dealt with later (see Chapter 3). The important point to recognize is that descriptive surveys chiefly tell us how many (what proportion of) members of a population have a certain opinion or characteristic or how often certain events occur together (that is, are associated with each other); they are not designed to 'explain' anything or to show causal relationships between one variable and another.

Descriptive surveys are well known and important. Any form of census falls into this category, as do many public-opinion polls and commercial investigations. Such surveys provide governments, manufacturers, economists and municipalities with information necessary for action. The job of such surveys is essentially fact-finding and descriptive — although the data collected are also often used to make *predictions*, for instance by comparing the results of similar surveys at different times and producing a trend, or by trying to forecast election outcomes or the number of homes that will be required in ten years' time and so on. Representativeness (or sometimes full enumeration) of a defined population is a first requirement.

There are many questions that actuarial surveys cannot answer or can answer only inadequately. Such questions usually start as 'why' questions and then proceed to examine group differences from which relationships between variables can be inferred. For instance, in comparing the results of election polls, we may find a rise, over a period of time, in the percentage of people choosing a particular party. We may wonder why this is happening and wish to explore the suggestion that, say, this party's policies with regard to old-age provisions makes it very attractive to older voters. We go back to our data and arrange the results according to age to see whether there are marked age-group differences in the expected direction. We could go further and try to find out whether low-income groups, or childless elderly couples, or young couples with elderly dependents are more prone to vote for this party, but clearly there is a limit to what we can do in this way: we may not have obtained the necessary information about income or elderly dependents; there may not be enough low-income voters above a certain age in our sample; we may not have asked any questions bearing on pensions and so forth.

We have now moved from questions about 'how many' to questions about 'why', from enumeration and description to an analysis of causality, and so we shall require a different type of survey, a survey with an *analytic* design.

Problems of causes and effects

Many current ideas about causality have their origin in what we think of as 'science' or the scientific method. All science is concerned with relationships or co-variation between variables, but some of our ideas about relationships or associations have to undergo subtle changes as we move from the physical and biological sciences to the social sciences. We may start with the *'monocausal' model*, suggesting that a single cause can have a specific effect. Thus, if we put a kettle of water on a fire it will get hotter and eventually it will boil; similarly, the harder it rains, the higher will be the level of a river and, eventually, it will overflow its banks. In the 'heroic' early years of microbiology the thinking among researchers was often monocausal: one disease, caused by one strain of bacteria, to be cured by one vaccine or other anti-body. Associated with these monocausal or 'A causes B' models were several other ideas. For example, causality of this kind was assumed to be linear, proportional and incremental: the stronger the cause, the bigger the effect. In many cases this type of causality was also expected to be reversible: if the cause were weakened or removed, the effect would decline and disappear.

However, the monocausal model often does not tell the full story, and sometimes it can be seriously misleading. For example, the level of the river Thames at London Bridge is not determined solely by the amount of rainfall in the Thames valley. It can also rise or fall with the tides, with the phases of the moon, with the state of the flood defences on the river banks and the position of locks and weirs up-river, and with the direction of strong winds from the

North Sea. Each of these causes will have a greater or lesser effect depending on its strength and timing and so we have to learn to think in *multi-causal* terms. Not only that, but the effects of these causes may be singular or *cumulative*: when adverse conditions of wind and tide, moon and rainfall combine at the same time, we may well get an exceptionally high water level and even flooding. Likewise, though the cholera bacillus is present in small numbers in most communities, it does not usually become a dangerous epidemic unless certain other conditions obtain: poor hygiene, contaminated water supply, lowered resistance among certain members of the population.

Let us also note that, in a multi-causal model, the several causal factors may be independent or *interrelated*. Thus, the twice-daily tides and the phases and diurnal orbits of the moon interact with each other, as well as with the water level in the Thames. In developing countries or under disaster conditions, poor hygiene, lowered resistance and water contamination often go hand-in-hand to create the conditions in which cholera flourishes.

Nor is it always necessary for these interactions among causes to happen simultaneously. Their accumulation may take many years and form a succession of influences, for example the silting up of a lake due to industrial effluent, followed by the proliferation of algae and the death of fish and other forms of marine life. Or the death of an individual from lung cancer may be due to a degree of genetic predisposition followed by a stressful lifestyle and prolonged heavy smoking. This is sometimes called a *causal pathway* or, in people, a 'career'. Thus, a young person does not become a delinquent in a monocausal way, for example, by coming from a 'broken' home. Other factors, successively or simultaneously, help to create a delinquent career: upbringing factors, educational factors, friends and the 'delinquent sub-culture' and so on.

When trying to disentangle problems of causality we often find *associations* or *correlations*, but of themselves these are no proof of causality. We might find, for example, that people who have acquired a microwave oven also have poor dietary habits, or that children with access to a computer at home are also better at mathematics in school. Does this mean that possession of a microwave or a home computer *caused* the dietary habits or the maths performance? On the basis of such associations alone we cannot say. It might be true that A caused B; or it might be that B caused A; that is, that people with poor eating habits are more likely to buy a microwave and that children who are good at maths try to get their parents to buy a home computer. Or it could be that another determinant, factor C, was the cause of both A and B. People with a busy lifestyle might have poor eating habits and also need a microwave; children from better-educated families might both be good at maths and have access to a home computer. We note, therefore, that associations may or may not denote causality; that such causality might run from A to B, from B to A or from C to both A and B. And often there is an interactive network of several causes.

Some mothers score higher than others on a measure of strictness. We may find that they are more often working class; that they have larger families, on average; that they have completed fewer years of full-time education; and that they are less likely to read women's-magazine articles about child upbringing. It would be quite possible to map out these associations and then speculate about causal interpretations. However, we shall also find that working-class women are less well educated, on average; that people with less education tend to read less; that in large families there is often neither the time to read nor the money to buy magazines containing articles about upbringing and so on. Each of the associations is not just linked independently with maternal strictness; it is also likely to be linked to the other associated factors and co-vary with them, as well as with strictness. It would take quite a complex research design and statistical analysis to disentangle these strands and to establish causal links from such a network of associations.

The nineteenth-century British statistician Sir Francis Galton once conducted an inquiry into the 'Objective Efficacy of Prayer'. He operationalized his inquiry by trying to find out whether clergymen lived longer than members of other professions. (So far, by making a few assumptions, he appears to have followed a monocausal model: more praying causes longer life.) His data did indeed show a small association: a difference of between one and two years in favour of the clergy. But at this point Galton seems to have adopted a multi-causal model of longevity. He expressed the thought that such a small difference in average age attained might well be accounted for by easier life circumstances, concluding that 'the prayers of the clergy for protection against the perils and dangers of the night, for protection during the day, and for recovery from sickness, appear to be futile'.

Strictly speaking, though, even if Galton's data had shown a very large difference in the average age attained by clergy compared to other professions, this would not have constituted proof of a causal connection between prayer and longevity. Some pre-existing, self-selective factor might well account for the results; for example, it might be that in those days the clergy attracted exceptionally healthy recruits with a predisposition to longevity. Or the results might be explained by another, more current, association; clergymen might be sustained into old age by exceptionally supportive wives!

We must also consider time, or rather temporal sequence, when seeking to establish causality. Our understanding of causality includes the strong presumption that causes must antecede their effects. Since it is implicit in this view that *earlier* causes are likely to be more important, this may lead us to discount later causes. Thus, some women suffer from post-natal depression; but was the baby's birth the cause of the depression, or was the woman suffering from episodes of depression long before the pregnancy? We must not assume that just because B followed A, therefore B was *caused* by A. The baby's birth may, or may not, have been the significant event that caused the mother's depression. We merely picked up this association because the two events happened at about the same time.

To establish causality with any certainty is difficult; to establish associations is often easier but leaves us to speculate about causality. It sometimes happens that association patterns can give us a strong hint about causality, and perhaps suggest an effective intervention method, even though the cause remains unknown. John Snow, the London physician who worked in Soho in the nineteenth century, was the first to show the connection between cholera and a contaminated water supply. He meticulously mapped the addresses of those who died from cholera during an epidemic at a time when micro-organisms were unknown and the beliefs about the causes of cholera were mainly superstitions.

His data showed that all those who had contracted the disease had taken their drinking water from the Broad Street pump. When he eventually persuaded the parish council to remove the handle from the pump, the epidemic stopped. All Snow had, by way of evidence, was an association; discovery of the cholera bacillus still lay in the future. However, disjoining the association had the desired effect — though it has been argued that the result was only partly causal because the epidemic was on the wane anyway!

Even today in medicine it is not unusual to find health measures being recommended for the cure or prevention of ills whose causes are as yet unknown. The expected efficacy of these measures rests on the disruption of patterns of association, for example between coronary heart attacks and various dietary factors. Sometimes, too, the reverse happens: we may find an effective remedy, even though we do not understand how it works, for example quinine and malaria.

The laboratory experiment, if properly conducted, is the classical tool for establishing causality. It does so partly by systematic manipulation of particular causal factors (to show changes in causal factors being accompanied by changes in effects) and partly by eliminating as many other factors as possible. Thus, pre-existing causes may be eliminated by allocating subjects *randomly* to the various experimental conditions; co-variation of causes may be explored systematically and interaction effects observed; and measures such as a double-blind design can reduce the risk that human experimental subjects will influence the results by their own expectations, or that the experimenters or their assistants will unwittingly convey their expectations to their subjects. But critics of human experimentation will often point to the difficulties experienced in creating truly experimental manipulation of psychological variables and to the fact that causes are not always what they seem: in survey research we can develop complex analytic designs to establish patterns of associations, but we can hardly ever prove causality because the experimental method is not usually open to us. However, we can sometimes participate in, or even help to plan, so-called natural experiments or social interventions or intervention programmes which, properly conducted, can permit us to draw causal inferences.

Taking part in social intervention programmes confronts us with the fact that most societies have strongly held world views and many systems of causal attribution. For example, many religions have built-in assumptions about the causes of poverty, ill-health, crime and depression, and other assumptions about what makes a good marriage or a happy childhood. Similarly, the authorities responsible for law and order hold strong assumptions about the causes of criminal acts, traffic accidents, football hooliganism or prostitution. Typically they will adopt a monocausal model, usually one embracing the personal responsibility of an individual. Thus, the responsibility for a car crash will be attributed to the driver with traces of a drug in his bloodstream. That the accident occurred on a notoriously dangerous bend, that the car crashed into a poorly sited telephone pole, that the road was made slippery by rain or ice, that the car's tyres were almost bald, that visibility was very poor at that time of night, that the driver was blinded by the lights of an oncoming vehicle are all facts that are often ignored, requiring a multi-causal model which the legal system would find difficult to handle. Societies often indulge in scapegoating: if

a prisoner escapes from gaol, the prison governor is sacked, although he probably had no knowledge of the prisoner and no direct influence on the events of the escape. This is done in the name of an attributional system called 'accountability'. If a child commits a dangerous or criminal act, we demand restitution from the parents although they were not directly involved; we do this in the name of 'responsibility'. Such attributional systems have little or nothing in common with valid cause-and-effect analysis, but simplistic mono-causal models and punitive attributions are easier to use than more appropriate but more complex causal models, and they are culturally sanctioned and widely available. More appropriate multi-causal models, on the other hand, would demand resources, time and skills for their validation and further research to examine the effectiveness of appropriate intervention programmes. Thus the social researcher who seeks to establish appropriate causal models and to validate them not only faces the daunting problems of research design but also encounters much societal resistance.

The future of causal attribution

Social research is trying to move from intuitive causal attributions (such as can be found in scapegoating, collective representations, stereotypes, health-belief systems, magic, folklore) to more objective kinds of causal attribution which are capable of verification and generalization. We have come part of the way with the aid of the research designs and strategies discussed above, but we still have a long way to go. This is partly because many of our ideas about research design and statistical analysis have come to us from other branches of knowledge, for example from the natural sciences, where other processes of causality predominate and where the linear approach to measurement has often been of good service. But societies and human beings are in important ways different from agricultural plots or biochemical reactions, and so we often feel that to apply some of these research designs to them is rather like putting our subjects on Procrustean beds.

For example, we rightly seek to distinguish between 'mere correlates' (associations) and 'actual causes', but the reason we cannot give associations the status of causes is because we do not know their place in a complex network of causality. The very word 'network' is a metaphor. It implies assumptions about the ways in which multi-causality works but, in fact, *we often cannot say what the correct causal model would be*. Take, for example, the long-term stability of belief systems and attitudes. Research shows that older people are often more religious than younger people and strongly believe in the hereafter. Does this mean that they had a more religious upbringing because they were educated long ago and have not changed since? Or have their religious ideas been subject to many influences and changes throughout their lives? Or have they recently 'found' religion because a strong belief system helped a close friend to come to terms with impending death? Our models of attitude measurement and change seem too linear to reflect such processes. Again, we often find that liking a school subject and being good at it go together, but by what means or metaphors can

we describe and measure how such an association develops? The same might apply to alcoholism and drink-related occupations; or to the stages of becoming a criminal. Such processes are ill-served by our traditional causal models.

We seem to have inherited notions of causality which imply a linear, additive progression, a kind of geological or 'layered' causality in which bounded one-way influences are brought to bear on people at successive times causing them to change. Our designs and our statistical techniques predicate us to think in this way because we know how to apply them; this kind of causality we can more or less handle. But in social research we often come across what might be called 'spiral reinforcement' processes. A small success at school, or a word of praise from a teacher, may make a child just that little more interested in a particular school subject, the child tries harder and develops its interest in that subject which in turn may lead to more praise and more success and so on. The process is incremental but not additive, influential but also interactive. The same may apply to the drinker who becomes a barman. A pre-existing liking for drink may lead by mutually reinforcing steps — a self-selective job choice, easy availability of drink, increased liking, membership of the drinking scene — to the development of problem drinking and so, ultimately, to dependence and alcoholism. As yet we have no models that can display, nor research designs that can reveal and predict, the workings of such causal processes.

In choosing our research strategy, our research design and statistical analysis we should therefore remain aware that these contain implicit assumptions about causal links and causal processes in people. Indeed, such assumptions may long ago have insinuated themselves into our own research interests and into our thought processes about causal attributions. While we can make good use of existing research methods in the service of replicability, data disaggregation and representativeness, we must not forget that human lives and human causality are not composed of layers of regression coefficients. We need a fundamental re think of the nature of the social influencing processes themselves. This may lead to the development of more appropriate analytic designs and to better models of human causality.

Selected readings

GENERAL TEXTS ON RESEARCH METHODOLOGY

Rossi, P.H., Wright, J.D. and Anderson, A.B. (eds), 1983, *Handbook of Survey Research*, Academic Press, New York.
> A good, comprehensive collection.

Kidder, Louise H., Judd, Charles M. and Smith, Eliot R., 1986, *Research Methods in Social Relations*, Holt, Rinehart and Winston, New York.
> Published by the SPSSI as a successor to the earlier text by Selltiz, Wrightsman, and Cook (see below), it is a comprehensive text at the undergraduate level.

Kidder, Louise H., Judd, Charles M. and Smith, Eliot R., 1986, *Research Methods in Social Relations*, Holt, Rinehart and Winston, New York.
> Published by the SPSSI as a successor to the earlier text by Selltiz, Wrightsman, and Cook (see below), it is a comprehensive text at the undergraduate level.

Moser, C.A. and Kalton, G., 1972, second edition, *Survey Methods in Social Investigation*, Heinemann, London.
> A well established, statistically orientated general text in survey methods, especially strong on sampling methods but also good on interviewing, questionnaire design and scaling methods.

Hoinville, Gerald and Jowell, Roger, 1978, *Survey Research Practice*, Gower, Aldershot, Hants.
> Straightforward and practical general text on survey methods. Especially good on organizing fieldwork and postal surveys.

Babbie, E., 1989, second edition, *Survey Research Methods*, Chapman & Hall, London.

Kerlinger, F.N., 1964, *Foundations of Behavioural Research*, Holt, New York.
> A comprehensive research methods text, especially strong on measurement and statistical problems.

Hippler, H.J., Schwarz, N. and Sudman, S. (eds), 1987, *Social Information Processing and Survey Methodology*, Springer Verlag, New York.

Maclean, M. and Genn, H., 1979, *Methodological Issues in Social Surveys*, Macmillan, New York.

Schuman, H. and Kalton, G., 1985, 'Survey methods' in Lindzey, G. and Aronson, E. (eds), *Handbook of Social Psychology*, Vol. I, third edition, Random House, New York.
> An extensive textbook chapter touching on all the main points of survey research.

Altreck, Pamela L. and Settle, Robert B., *The Survey Research Handbook*, Richard Irwin Inc., Homewood Ill.
> An easy to read undergraduate text.

Singer, Eleanor and Presser, Stanley, 1989, *Survey Research Methods*, University of Chicago Press, Chicago.
> A reader dealing with sampling, non-response problems, validity, interviewer bias, telephone interviewing and related topics.

Przeworski, Adam and Teune, Henry, *The Logic of Comparative Social Enquiry*, Wiley (Interscience), Chichester, Sussex.
> Deals with the logic of comparative studies and the problems of equivalence of measures. For advanced students.

Sage University Papers Series: Quantitative Applications in the Social Sciences.
> An expanding series of small paperback volumes dealing with every aspect of social research methods.

The Open University, Course DEH 313, 1993, *Principles of Social and Educational Research*, The Open University, Milton Keynes.
> An excellent set of self-study course books, available separately and supported by exercises and by collected readings.

Fink, Arelene and Kosecoff, Jacqueline, 1985, *How to Conduct Surveys*, Sage, London.
> A text for beginners.

Selltiz, C., Wrightsman, L.S. and Cook, S.W., 1976, third edition, *Research Methods in Social Relations*, Holt, Rinehart and Winston, New York.
   Still generally useful.

Hyman, Herbert, 1955, *Survey Design and Analysis*, Free Press, Glencoe, Ill.
   A major older research text full of useful applications. Helpful on descriptive versus analytic designs.

Bynner, John and Stribley, Keith M. (eds), 1978, *Social Research: Principles and Procedures*, Longman and The Open University Press.
   A useful collection of original papers on research design and methods of data collection.

ON PUBLIC OPINION POLLS

Bradburn, Norman M. and Sudman, Seymour, 1988, *Polls and Surveys*, Jossey-Bass, London.

2
ANALYTIC SURVEY DESIGNS

The analytic, relational survey is set up specifically to explore the associations between particular variables. Its design is in many ways similar to that of the laboratory experiment. However, like experiments in the laboratory, it is usually set up to explore specific hypotheses. It is less orientated towards representativeness and more towards finding associations and explanations, less towards description and enumeration and more towards prediction, less likely to ask 'how many' or 'how often' than 'why' and 'what goes with what'.

Four types of variables

In designing an analytic survey, it is helpful to distinguish four different kinds of variables:

1. *Experimental variables.* These are the 'causes' or predictors, the effects of which are being studied. They are sometimes referred to as 'independent' or 'explanatory' variables. The analytic type of survey, like the lab experiment, is set up to vary these factors systematically so that their effects can be observed. Often several such variables working both in isolation and in various combinations are of interest.
2. *Dependent variables.* These are the results, the effects–variables, the gains or losses produced by the impact of the experimental variables, the predicted outcomes. These variables have to be measured particularly carefully and group differences tested for statistical significance.
3. *Controlled variables.* As a source of variation these should be eliminated in order to fulfil the condition of 'other things being equal' when the effects or correlates of the experimental variables are stated. Variables can be controlled by *exclusion* (for example, by having only males in a sample, gender is excluded as a source of variation); by *holding them constant* (for instance, by interviewing all respondents on the same day, thus eliminating day-of-the-week effects); or by *randomization* (for instance, in the case of a multiple-choice question, by systematically randomizing the order in which the alternatives are presented to the respondents, thus eliminating ordinal and serial effects as a source of variation).
   Uncontrolled variables. These are 'free-floating' variables and can theoretically

be of two kinds: (a) *confounded variables* and (b) *error*. The confounded variables, sometimes called 'correlated biases', have hidden influences of unknown size on the results. For example, in medical research the results of treatment may be affected by a hitherto unknown allergy or by unsuspected side effects; in some psychological research, genetic influences may play a much bigger part than anticipated; in advertising, public reaction to a new product or campaign may be unexpectedly volatile. Essentially this means that knowledge and understanding of the phenomena under investigation are still incomplete in important ways; there are variables, other than the experimental and controlled ones but confounded with them, that can affect the results and hence can produce serious misinterpretations. On the other hand, such uncontrolled variables can lead to the development of new hypotheses so that, eventually, their impact may be controlled.

Inevitably, any research design also suffers from error. Such error variables are (or are assumed to be) randomly distributed or, at any rate, distributed in such a way as not to affect the results.

In practice it is not usually possible to distinguish between confounded variables (that is, hidden additional causes) and 'pure error'. In analytic surveys, as in experiments, the influence of uncontrolled variables is made as small as possible. If the presence of confounded variables is suspected, a good deal of statistical analysis may be required to uncover their identity, and ultimately their impact must be studied systematically in a new enquiry.

An example may help to clarify these terms. Let us assume that we are trying to understand some of the determinants of children's bedtimes. We have decided to study the effects of age, which will be our experimental variable, and the survey will obtain information from a sample of children for every day of the week. Our dependent variable will be the children's bedtime, in all its variations. There are many variables that we will have to control in order to observe the effects of age. For instance, children who have recently been ill may have to go to bed especially early; we will probably control this variable by excluding such children from the sample. Children go to bed later in the summer and during school holidays; we may control these variables by holding them constant, by collecting all our data during one short period in the school term. Children go to bed later if they have older brothers or sisters; we can try to take care of this by making sure that children with older siblings are randomly distributed through our sample.

There remain a considerable number of uncontrolled variables, some of which are very likely to be confounded. For instance, socioeconomic background is likely to be an important influence on bedtimes; if we realize this in advance, we may be able to control it, but otherwise this factor can easily introduce a bias in our conclusions. It may, for instance, happen that the older children in our sample also come from more well-to-do homes, where bedtimes tend to be earlier; in that case, we may wrongly conclude that increasing age is *not* a good predictor of later bedtimes because, unknown to us, the socioeconomic factor is counteracting the age factor. Again, the child's gender, membership in a youth organization, or a keen liking for television may be important determinants of bedtimes in parts of our sample. Such uncontrolled variables, unless they are known to be randomly distributed, can bias the results to an unknown degree. The 'ideal' design would contain no uncontrolled variables.

For our design to be effective, each of the four types of variable must be measured as carefully as possible. This is of particular importance with regard to our dependent variable(s). It often happens that, in designing a study, people become so involved with the problems of the experimental variables (the 'causes') or the controlled variables, that they give insufficient attention to the dependent variable — or even forget to measure it altogether! Thus, in the earlier example about children's bedtimes, they may simply forget to ask any questions about bedtimes — without which all the other questions are useless. It will also soon become clear that 'bedtime' is not a precise concept: does it refer to the 'official' bedtime (if any) laid down by the parents, to the time the child actually gets into bed, to the time the child falls asleep? Do we ask the child, do we ask the parents or perhaps an older sibling, if any?

In deciding how to measure such a variable, we first have to define it and then define it further in operational terms. This process of defining should take us right back to the initial aim or purpose of our research. If we are, say, conducting some kind of medical enquiry about the need for rest and sleep in children, then 'bedtime' might well be defined as 'falling-asleep time' (to be followed by 'waking-up time' and questions about interrupted sleep). But if we are interested in, say, social-class differences in parental standards, then 'bedtime' might be more appropriately defined as the 'official' bedtime (if there is one). Thus, the definition of a variable will not depend on a dictionary but on the internal logic of the inquiry.

Further pilot work will then show how a variable, as defined for our purposes, can best be measured in the field. The pilot work may even show that we need still further refinement in our definition. We should never shirk any effort to make our dependent variable more precise and more robust. If we had simply asked a sample of parents about their children's usual bedtimes, then it would be very difficult to draw conclusions about differences in age, gender, school term versus vacation, week-ends etc. because parents may have interpreted our question in several different ways — with such an ill-defined dependent variable the rest of the study becomes almost meaningless.

Often there are difficulties in trying to define and measure a dependent variable. Imagine, for example, a study of the determinants of recovery from alcoholism; the term 'alcoholism' is difficult enough in itself, but to define 'recovery' from it is more difficult still. Or suppose you have been asked to measure the effects of a health education campaign, say with regard to the prevention of coronary heart disease. What exactly would be your dependent variable? A measure of knowledge? A measure of 'acceptance' or 'awareness' of the campaign's message? A set of cause-of-death statistics? Data from all the local hospitals?

Or again, suppose you have been called in by a municipal transport authority to find out whether the painting of bus lanes on certain busy roads (into which cars are not allowed between certain hours) is doing any good. What would be your dependent variable(s)? Average bus speeds before and after? Accident rates? Delays to motorists? Police expenditure? How, furthermore, would you try to measure each of these variables free of bias? And how would you make

allowance for different times of day, different days of the week, different classes of road users, inbound versus outbound traffic flows and the novelty impact of such an intervention?

Cross-sectional designs

Earlier we compared the bedtimes of different children at different ages, on the assumption that these comparisons could tell us what would happen to children over a period of time. We did not actually take the *same* group of children and follow them over a number of years. We hoped that the groups of children were comparable and that we could thus in fact observe the relationship between age and bedtimes. This type of design is known as cross-sectional to distinguish it from longitudinal or 'before-and-after' designs.

Suppose that in our example we had obtained a probability sample of children, say, between the ages of five and fifteen. What often happens is that such a sample is collected for a descriptive or actuarial purpose and that a kind of cross-sectional design is imposed on it afterward. First, we obtain the over-all average bedtime. Next, we compare the different age groups and find that bedtimes get later with increasing age. Then, we decide to compare the sexes, and we find that boys, on average, go to bed a little later than girls. We go on in this way, comparing bedtimes against a number of variables in turn such as social class, size of family, urban/rural differences and so on. Suppose that each time we find a difference in bedtimes. We now want to go further; we want to know, for instance, whether the age differences still remain if we hold social class constant. This means that for each of our social-class groupings, we have to compare the bedtimes of children in their different age groups separately. We could go still further and try to study the effects of three variables in combination, say, social class, urban/rural and size-of-family differences. However, to do this we need to have a sufficient number of cases in each of our combinations or 'cells' (unless our predictions refer only to *part* of the sample). Some of these cells will be readily filled, for instance those for children from large families in urban districts with working-class backgrounds. Other cells will be difficult to fill or may remain empty, such as those for children from small working-class families in rural areas. This is because a sample selected to be representative of the child population as a whole will not contain sufficient numbers of these rarer combinations to make comparisons possible.

We have moved to a different kind of research problem. Now we no longer ask 'How many?' but 'What goes with what, and why?' We have moved from a descriptive to an analytic type of survey that tries to answer questions about relationships and determinants, and therefore we shall need a different type of sample, a sample that is primarily geared to making comparisons.

Upon re-examining the preceding example, we note that our difficulties arose partly from running out of cases to fill the cells representing rare combinations and partly from the fact that some of our experimental variables were related to one another as well as to the dependent variable (bedtimes). Thus, in finding differences between children with urban or rural backgrounds, we wondered whether this difference would disappear if we held social class constant because

the proportion of working-class children was much higher among the urban part of our sample. Perhaps, if there had been as many middle-class children as working-class children in both the urban and the rural parts of our sample, the urban/rural differences would have been negligible.

Another point we note is that it becomes necessary to *plan in advance* the comparisons we wish to study and to make them part of our research design; we cannot usually apply this kind of analysis to a sample which has been collected for some other purpose.

Factorial designs

The problem of interrelated independent variables has led to the development of factorial designs, whose particular function is to disentangle complex sets of interrelationships. They are analogous to the 'agricultural plots' type of design which led to the development of analysis-of-variance techniques. Analysis of variance is, however, chiefly applicable to quantitative variables, whereas surveys deal mostly with categorical variables (see Chapter 9).

This type of analytic survey design is one of the ways in which we can approximate laboratory conditions. However, in laboratory experimentation we create or introduce our experimental variables and, while controlling for most other variables, observe the concomitant changes in the dependent variable. In survey research we are not normally in a position to impose experimental factors or to manipulate the lives of our respondents, for instance by allocating them at random to an experimental or a control group. Instead, we select respondents who already have the characteristics required by our design and compare them in their groupings. This makes it impossible to give causal interpretation to any pattern of associations we may find. Instead of dealing with just one experimental variable and controlling all others, the factorial design enables us to study several experimental variables *in combination*. Not only does this provide us with more information but also with greater confidence in predicting the results under various circumstances.

At the outset, we have to decide which are going to be our experimental variables — we will want to vary these systematically and in combinations with each other. We must aim to control all other variables by exclusion, holding constant or randomizing. The choice of experimental variables predetermines what we can hope to get out of our study; if we decide, later on, that some other variable should have been chosen, it may not be possible to show its effects except by selecting a new sample. This means that we must have an adequate general idea beforehand of the lay of the land in our particular area either through previous studies or from the pilot work.

The choice of our experimental variables will be limited by practical considerations. We might start off rather ambitiously by choosing, say, sex, age, class, urban/rural and size of family in our bedtime study. Let us assume that we have ten age divisions, seven socioeconomic grades and five family-size groupings, while sex and urban/rural are dichotomies. This would give us a design containing $10 \times 7 \times 5 \times 2 \times 2 = 1,400$ cells, which, even at no more than ten cases per cell, would require 14,000 children. If this is unacceptable, we can

either cut down the number of experimental variables, or make them cruder (for example, by reducing class from a seven-point scale to a dichotomy), or probably both. In this way we can reduce the number of cells and, thus, the number of cases required, to more manageable proportions. Even so, a design requiring three dozen cells and forty cases per cell would not be at all unusual. Equal (or proportionate) numbers of cases are required in each cell in order to disentangle the experimental variables from each other. Respondents in each cell should be a probability sample (Chapter 3) of all those eligible to be in that cell.

A disadvantage of the factorial design is the need for careful selection of cases. Sometimes this will mean that a special *enumeration stage*, using only a short questionnaire, will have to precede the main inquiry (the much larger enumeration sample being used to supply the respondents to fit each cell of the factorial design). At other times, the number of respondents in some cells will unavoidably remain below requirements, and in that case, some method of statistical extrapolation or 'randomized replication' will have to be used.

The factorial design makes no pretence of being representative. Indeed, in its final form it is often markedly unrepresentative, owing to the need to incorporate equal or proportionate numbers of respondents in each cell, so that very rare combinations of determinants may be encountered as often as the more popular or typical ones. (However, it is sometimes possible to extrapolate and adjust the figures afterwards, using 'correction factors', to make this kind of sample yield population estimates also.)

The factorial design permits us to vary our experimental variables systematically and in combinations of two or more, though such patterns of multiple association cannot be interpreted as causal. We also need to introduce measures to control the remaining variables and to eliminate as many uncontrolled variables as possible. Some of this can be done, as before, by exclusion, holding constant and randomization, but inevitably some confounded variables may remain. In the later stages of the analysis, various hypotheses concerning these confounded variables will be explored by means of special cross-tabulations, matching of subsamples (see below), and by multivariate analysis. While a confounded variable may turn out to have an unexpected effect on the results, upsetting the regular pattern of the design as planned, it can also be stimulating to further research and the discovery of new relationships.

Designs using regression and other multivariate analyses

As we have seen, in social research we rarely deal with monocausal phenomena, that is with a single cause having a specific effect. Almost invariably we have to deal with multi-causal models, so that any effect is the outcome not of one cause but of a complex network of determinants. Quite possibly many of these will not only be related to the dependent variable but also to each other: they will form a network of *interrelated determinants*.

Suppose that we are interested in some aspect of school achievement: why do some children do better than others at, say, arithmetic? We shall soon find that this is not merely because some children are more intelligent than others, but because their performance is also related to several aspects of teaching at school,

to parental support with homework, to size of school class and so on; it is easy to think of two dozen or more possible determinants — far more than a factorial design could cope with. Each of these variables contributes to the outcome (the dependent variable 'achievement in arithmetic', as tested) to a different degree and in a different way. Also, we shall find that each determinant (experimental or independent variable) may be related to others; there may be more effective teaching in smaller classes or more parental support for more intelligent children. This makes it difficult to find out what influence each independent variable, of itself, has on the outcome or what would be the effect of changing a particular independent variable.

There are advanced statistical procedures, such as the various kinds of multivariate analysis (for example multiple regression analysis, path analysis), which are capable of disentangling this kind of associational network. Ideally, they can do three things for us. First, they can give us a firm idea of the total variance accounted for by our independent variables (that is, how much of the variation in our dependent variable we can 'explain' with the aid of the independent variables we have chosen and how much variation remains unexplained). Second, they can tell us which are the most important determinants and which are less important or insignificant. Third, they can tell us how powerful each determinant is after its links with other variables have been discounted (for example, do brighter children still do better at arithmetic, even when we 'allow' for parental help with homework?). In other words, they can tell us how powerful each determinant is, 'other things being equal', or 'holding everything else constant'. Sometimes, too, these statistical techniques can show what additional influence or strength each successive independent variable has added during the emergence of a causal network or what its unique explanatory power is.

These multivariate techniques are not always applicable. They rest on statistical assumptions which may be hard to fulfil and generally they require interval-type variables; that is, quantitative integers such as scale scores. Nominal or categorical data such as eye colour or marital status are more difficult to handle (see Chapter 9).

Moreover, even these more powerful techniques do not, strictly speaking, allow us to make causal references. Rather, they help us to disaggregate the variance (the fluctuations or differences) in the dependent variable according to the *relative* importance of the independent variables which we have entered into the design and to do this by statistical rather than by experimental means.

Multivariat analysis is not for the beginner. You will need access to a reasonably powerful computer, to the relevant computer software and to sound statistical advice. These techniques are best applied in well-researched domains such as education, in which it is possible to generate precise hypotheses and potential causal models about interrelationships between variables as well as adequate measures by which to test them.

One technique of multivariate analysis is called *multiple regression*. What will this technique do for you? First, having listed all your hypotheses and constructed your measures of the dependent variable(s) and of, perhaps, several dozens of (potentially interrelated) independent variables, you will need to obtain your data. You will not need a preliminary enumeration survey such as

might precede a factorial design study; but you will need substantial numbers (or perhaps even a national sample of the relevant groups) to obtain adequate variability on your measures and to enable you to show that any group differences you may find have not arisen by chance (that is, are statistically significant), and are large enough to be socially relevant and useful. After processing and analysing the data the computer program will produce a table which lists all your independent variables one by one and will give a figure for each such variable to indicate how much it contributes to the 'explanation' of your dependent variable. A total will also be given, showing how much of the variance is explained by all your independent variables put together. If your measures and hypotheses have been strong, then together they might account for, say, 80 or 85 per cent of the total variance. In the social sciences we are more likely to have poor predictors, so a total of 40 or 50 per cent is more probable — leaving about half the variance unaccounted for. Left to its own devices, the program will produce this first table showing your independent variables in order of the size of their explanatory contribution. Thus, this first pass through the data will give you two things: (a) the total proportion of the variance accounted for; and (b) a list of coefficients, starting with a handful of significant ones and ending with a larger list of variables which have virtually no 'power', that is which at first sight contribute little or nothing to an 'explanation' of the dependent variable.

You may, however, wish to go further. An advantage of multiple regression is that you, the researcher, can determine the order or sequence in which your variables will enter the analysis. You may have testable hypotheses about the relative importance of each independent variable and feed them into the analysis in this order. Or you may wish to test a general model. For example, in educational research you might wish to determine the order of entry of your independent variables according to the developmental stages in a child's life: first, the pupil variables (such as IQ and quality of the home); then, the school-related variables (such as average expenditure per pupil); after that perhaps the teacher-related variables (for example, proportion of teachers with relevant specialist training), and finally the classroom-related variables (such as size of class, availability of equipment) — all depending on how the research design has been set up and what questions it is trying to answer.

While there are advantages in this flexibility, a word of caution is needed. This is because many of your variables are likely to be interrelated, and so the impact of any given variable may be shared by that of others. This makes the results more difficult to interpret. It all depends on the type of associational network you are studying and on the degree of independence of your variables.

This discussion of regression analysis has taken us rather far afield, yet it is only one of a number of multivariate analysis methods. We might, instead, have chosen to use analysis of covariance, factor analysis or some form of discriminant function analysis. These all require more advanced statistical knowledge (as well as the availability of appropriate computer software), and your choice will depend on the kind of problem you are investigating. For example, cluster analysis or discriminant function analysis are probably most appropriate when you are seeking to develop a typology, a way of classifying your subjects into a few distinctive sub-groups. Factor analysis (see Chapter 9),

on the other hand, would be appropriate when you are trying to discover the main underlying determinants of your data or to test a set of hypotheses about them.

In choosing the appropriate statistical technique much will depend on previous research, on how well understood the particular domain of enquiry is. In some fields we have a number of reliable quantitative measures, some precise theories and some useful causal models and so we may be able to apply multivariate techniques of analysis with advantage. In other fields we would hardly know where to start, and the use of such analytic techniques would be premature. What would we do, for example, if we were asked to find out why some people become amateur clarinettists or start to invest in the stock market, while others do not?

$STOP$ ———

Before-and-after designs

We have noted repeatedly that cross-sectional, factorial and multivariate types of design cannot determine true cause-and-effect relationships; they can only give us information about associations or correlates — though some of these are very suggestive! *Before-and-after designs* (sometimes referred to as pre-test/post-test designs) have been developed in an effort to overcome this disadvantage.

The form of the simplest before-and-after design is indicated by its name: a set of measurements (base-line measures) is taken of a group of respondents, who are then subjected to an experimental variable and afterwards measured again — once more or perhaps several times. The difference between post-test and pre-test results or observations is said to be the 'effect' of the experimental variable, though this is misleading.

At first sight it may appear that this design will present far fewer problems in respect of controlled and uncontrolled variables and error because the respondents in the experimental group 'act as their own controls'. They provide the 'other things being equal' conditions that enable us to isolate the effects of the experimental variable on the dependent variable so that we can draw valid causal conclusions about the process.

But can we? We cannot legitimately attribute all the before-and-after differences to the effects of the experimental variable we are investigating until we are sure that, without it, such changes would not have occurred or would have been smaller or different. Depending on the time interval between pre-test and post-test (a few hours, a day or two or perhaps many months), it is always possible that some changes in the expected direction may take place even without the impact of the experimental variable. For example, some of the people whom we have subjected to an advertising campaign to buy a certain product or brand might have done so anyway; some patients who have been given an experimental treatment might have recovered even without this intervention. This may be due either to concurrent influences (for example a price reduction) or to pre-existing ones (for example a strong immune system). Moreover, if people are aware that they are participating in a survey or an experiment, this in itself will often produce certain changes — they become more aware or alert, they develop expectations about the outcome, and with all