3

Formulating and clarifying research questions

Research questions need to be focused. It is not enough to say, for example, 'I'm interested in getting some answers about inequality'. What answers to what questions? Do you want to know the extent of inequality, its distribution, its causes, its effects or what? What sort of inequality are you interested in? Over what period? Where?

The purpose of this chapter is to provide guidelines on how to clarify research questions. It will do this by:

1 outlining the difference between explanatory and descriptive research;
2 providing guidelines on how to focus descriptive research questions;
3 distinguishing between types of explanatory research questions;
4 identifying resources that help in reviewing existing research and thus help focus what research needs to be conducted; and
5 showing how research design is a key element in shaping the way research questions are asked.

The chapter stresses the need to formulate a research question at the beginning of the research—certainly before data are collected. But this emphasis does not mean that the initial research question is the final one. While it is important to know what we are looking for, it is a mistake to let this initial focus blind us from other unanticipated questions which are perhaps more interesting, important or manageable than the initial one. Questions can be refined and new issues emerge while reviewing literature or collecting and analysing data.

TYPES OF RESEARCH QUESTIONS

There is no simple way to refine research questions but I will outline a number of different types of questions and provide some guidelines to help focus research. First, however, it is helpful to define a few terms.

A *variable* is a characteristic that has more than one category (or value). Thus sex is a variable with the categories male and female. Age is a variable with many different categories (one year old, two years old etc). Any case, however, will belong to only one category. A variable then is a characteristic on which cases can differ from one another. In cause-and-effect terms we can distinguish between three types of variables: *dependent*, *independent* and *intervening* variables (see Box 3.1).

The following diagram illustrates the three types of variables. Education is the independent variable that affects income level (the dependent variable) via its effect on job (the intervening variable).

> **BOX 3.1** *Dependent, independent and intervening variables*
>
> | Type of variable | Description Synonyms | Symbol | Causal diagram |
> | --- | --- | --- | --- |
> | Dependent | Effect
Outcome | Y | $X \rightarrow Z \rightarrow \mathbf{Y}$ |
> | Independent | Cause
Predictor | X | $\mathbf{X} \rightarrow Z \rightarrow Y$ |
> | Intervening | Mechanism | Z | $X \rightarrow \mathbf{Z} \rightarrow Y$ |

> **WEB POINTER 3.1** *Web reading on types of research and research questions*
>
> | | |
> | --- | --- |
> | Useful explanation of descriptive and explanatory research. | http://trochim.human.cornell.edu/kb/resques.htm |
> | Types of research questions. | http://trochim.human.cornell.edu/kb/resques.htm |
> | Brief discussion of variables. | http://trochim.human.cornell.edu/kb/variable.htm |
> | Brief outline of some ways of defining a research question. | http://trochim.human.cornell.edu/kb/probform.htm |
>
> Visit www.social-research.org to use these links and to check for updates and additions.

education → job → income
X   Z   Y
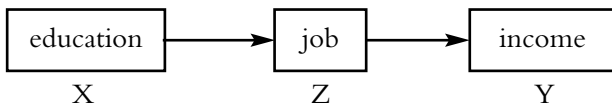
Once we have selected a broad topic ask 'what do I want to know about this topic?' Broadly, interests will be descriptive, explanatory or both.

Descriptive research

It is difficult to provide precise rules about how to focus a descriptive research question but five questions can help.

1 What is the *time frame* of our interest?
2 What is the *geographical location* of our interest?
3 Is our interest in broad description or in *comparing* and *specifying* patterns for subgroups?
4 What *aspect* of the topic are we interested in?
5 How *abstract* is our interest?

The way in which these questions help focus a broad research question is illustrated in Box 3.2.

Explanation: searching for causes or consequences

The next three types of research question all involve explanatory research. They only vary their focus and complexity.

Over the last 50 years the divorce rate in most western countries has increased markedly. We may want to know why but have no idea. Diagrammatically then the problem is:

? → increase in divorce rate
X   Y
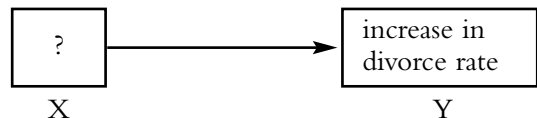
Alternatively our interest might be to discover the consequences of the increased divorce rate. Diagrammatically this is:

increase in divorce rate → ?
X   Y
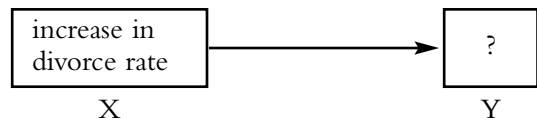
Start with a broad topic such as divorce. The following questions can help focus the topic.

1   *Time frame:* Do we want to know about divorce now or in the past, or do we want to look at the trends over, say, the last 50 years?

2   *Geographical location:* Do we want to know about divorce in a part of the country, for the whole nation or for other countries? Or is our interest *comparative*: do we want to compare one country with other countries? If so, which countries?

3   *Comparing and specifying subgroup patterns:* Do we simply want to know, for example, the national divorce rate or do we want to see if it differs according to age of marriage, occupation, length of marriage, educational level and so on?

4   *Aspect:* Does our interest in divorce centre on the divorce rate, divorce laws, problems with property and custody, attitudes to divorce, ways in which people adjust to divorce or something else?

5   *Abstraction:* Are we interested in the raw facts or in what they might indicate at the more abstract level? For example, is our interest in the divorce rate as such or in what it might reveal about, say, level or social conflict? If it is the latter, we will collect other data relevant to measuring conflict (e.g. level of industrial disputes, crime rates, legal actions). If it is in divorce *per se*, we will collect only information on this topic.

The first step in focusing an explanatory research question is to decide whether we are looking for causes or consequences. Second, we must clarify what it is we are seeking the causes or consequences of (e.g. increased divorce rate from 1950 to 2000). The next step is to list possible causes or conse-

quences and then collect relevant data. There are a number of ways of coming up with such a list.

1   *Previous research:* Using a library, the internet or other similar locations, look to see what other researchers have found on the topic (see Web Pointer 3.2). We will often notice gaps in the research, lack of evidence and unresolved debates. These can help focus the research.

2   *The 'facts':* Detailed description of a phenomenon can stimulate ideas about possible causes. For example, if the pattern of divorce over the last 50 years showed a sharp increase in, say, 1976 we could ask what other changes have also occurred at these times and over the 50-year period. These might have something to do with the divorce rate and provide clues about possible causes (e.g. decline of religion except in the mid-1950s, changes in the law in 1976, changed beliefs about personal fulfilment, increased workforce participation for women, changes in welfare provisions).

3   *Our own hunches:* Our own ideas, impressions, beliefs and experiences are valuable sources of ideas so long as we test them against the evidence and are not limited only to them.

4   *Talk to informants:* Often there are people who are particularly well placed to provide ideas and can provide good starting points for research. In this case marriage counsellors, solicitors and social workers may all provide helpful insights.

Once ideas about possible causes have been collected in this exploratory way we might stop there, collect information to test all the ideas or focus on just one or two possible causes.

Explanation: exploring a simple idea

This approach to formulating explanatory research questions is more focused than the previous approach. Instead of asking what has caused the increased divorce rate, we would be more specific and ask: 'has X (e.g. decline in religion) led to the increased divorce rate?' The research then concentrates on this specific question. Diagrammatically we could represent this as:

decline in religion	$\xrightarrow{\text{?}}$	increase in divorce
X		Y
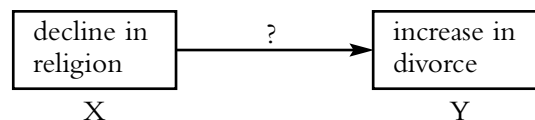
Explanation: exploring more complex ideas

The above idea is relatively simple because it deals with only two concepts. We might find that it is true that the decline of religion has been accompanied by an increase in divorce but what are the mechanisms? Why should religious decline lead to divorce? Can we specify the links? If we can, we should have a better understanding of why divorce has increased. From a variety of sources (see the earlier section on searching for causes and consequences) we might develop the model in Figure 3.1.

In summary there is a variety of types of research questions varying from descriptive to more complex explanatory problems. For descriptive research we must try to be clear and specific about what we want to describe. Box 3.3 lists four key questions to help focus explanatory research questions.

It can be helpful to draw diagrams like those earlier in this section: they can help clarify our thinking. Often these different types of research questions can represent different stages within the one study. Initially we might be interested simply in describing a phenomenon, but then the 'why' questions about causes develop. As we proceed we might focus on one particular cause and then start asking more complex questions about that particular factor. Remember, the task of clarifying research questions is a process rather than a once-and-for-all task which is completed at the beginning of a study.

> **BOX 3.3**
> *Four questions to help focus explanatory research questions*
>
> 1   What am I trying to explain?
> 2   What are the possible causes?
> 3   Which causes will I explore?
> 4   What are the possible mechanisms?

USING THE INTERNET TO REVIEW EXISTING INFORMATION AND RESEARCH

The process of focusing a research question requires a knowledge of the field, an understanding of previous research, an awareness of research gaps and knowledge of how other research in the area has been conducted. (See the checklist at the end of this chapter for some of the things that your review of the existing literature should examine.) The task of locating previous research on a topic has been made much simpler and faster with widespread access to the internet and the availability of electronic data bases, electronic journals, online journals and sophisticated search engines. The internet is useful in providing access to many types of information (see Web Pointer 3.2).
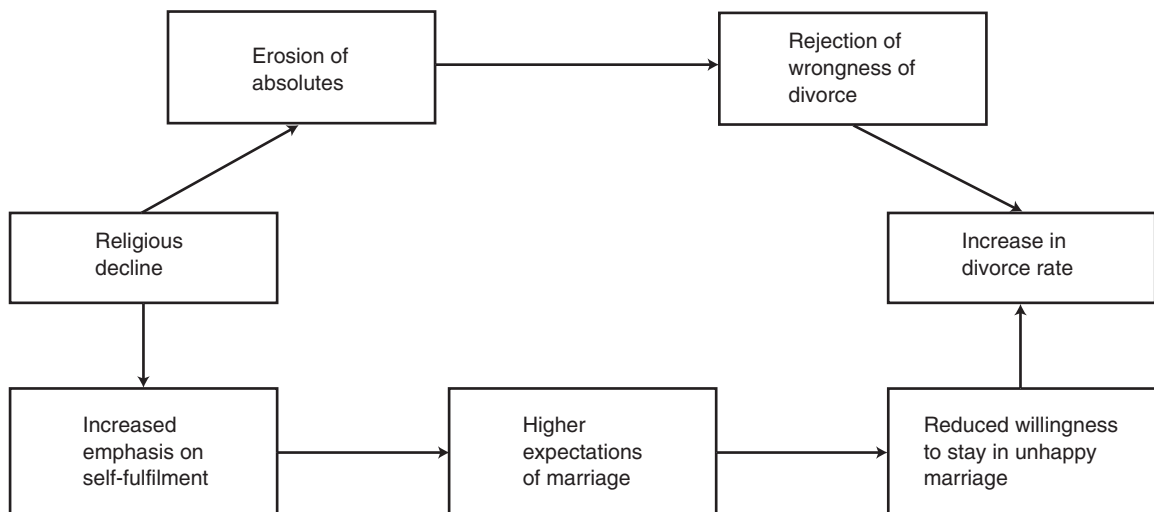

Figure 3.1   A model for the increased divorce rate

Before using this information you need to know how to use the internet effectively. You need to know how and where to look and to evaluate the information obtained. Using the internet there are seven main ways of obtaining information to assist in reviewing a research field.

1 *Information freely available to anyone with access to the internet:* There are three valuable strategies to locate such information.

   a *Use search engines:* Many search engines can locate content on the internet but no one engine searches the whole internet and some are better than others for particular types of material. There are two main types of search engine: the *single search engine* and the *metasearch* engine. Metasearch engines run a number of the single search engines at once and are therefore the best way of locating content.

   b *Follow links:* Many sites have links to other related websites. Clicking on these links will redirect you to that related site. In some ways the process is similar to looking at a bibliography at the end of an article and looking up the references that are listed.

   c *Use gateways:* Gateways are internet sites the purpose of which is to provide many links to other sites on a specified topic. These sites are directories in the same way that the *Yellow Pages* telephone directory lists businesses according to their service or product.

2 *Free access to online catalogues:* Many libraries provide easy, web-based access to their catalogues. Searching these catalogues can indicate the types of material available on given topics. Users with borrowing privileges can borrow these or order them on interlibrary loan. These catalogues usually do not provide indexes of journal articles (see points 5 and 6 below).

3 *Free access to publications and full text articles:* Free access to a large number of online newspapers is available. Many of these only provide access to current news items while some enable searches of older items. Some will notify users of news items on particular topics that appear in specified newspapers (news clipping services).

4 *Restricted access to resources such as online publications:* Many publishers are providing their publications electronically for a fee. Some free online publications (e.g. many newspapers) are supported by advertisements.

5 *Restricted access to lists of articles in academic journals:* These sites provide notification services of articles from nominated journals as they are published. Most of these services are provided only to registered and paid-up subscribers or libraries.

6 *Restricted access to searchable databases*: These databases contain details of academic articles published in a wide range of journals, from conferences, theses and some books. They provide bibliographic details and abstracts of each article and search engines enable focused searches. These databases are normally accessed through academic libraries. They include papers going back some years and are regularly updated. Databases exist for broad disciplines (e.g. sociology, psychology), for particular topics (e.g. AIDS research) or topics for particular countries (e.g. family research in Australia).

7 *Newsgroups, listservs and chatgroups:* These internet facilities provide a means of obtaining information and ideas more or less interactively. They allow you to ask questions of other people on the internet who visit these interactive sites. There are three main types of sites that are helpful for obtaining this type of information or for testing out ideas.

   a *Newsgroups* are publicly accessible topic areas to which anyone can subscribe without charge and receive and send messages. Some can be useful but many are burdened by junk.

   b *Listservs* are email mailing lists which internet users can join. Each list is dedicated to particular topics. Once a user subscribes to a list they automatically receive 'postings' to the list from other members and can send messages to other members of the list. They are a good way of sharing information and are generally much better than Newsgroups.

   c *Chat groups* enable subscribers to converse in writing with one another in real time. At any one time a number of subscribers of a particular topic-based chat group may be logged onto the internet-based chat group thus enabling those who are logged on to exchange ideas there and then.

WEB POINTER 3.2 *Using the internet to review existing information and research*

Provides a tour of some key internet sites for psychologists, sociologists and many other disciplines. Provides tools and techniques to improve your internet searching. Emphasises critical thinking when using internet resources. This is *the* place to start if you want to become good at using the internet for researching in your discipline area.	www.vts.rdn.ac.uk/
How to locate information on the internet.	http://library.albany.edu/internet/research.html
Excellent tutorial to help evaluate information on the internet.	http://tramss.data-archive.ac.uk/Research Questions/index.asp

Search engines

Describes how to use search engines.	http://library.albany.edu/internet
A brief guide to the effective use of search engines.	www.jisc.ac.uk/subject/socsci/strategy.pdf
Checklist for searching the internet.	www.jisc.ac.uk/subject/socsci/check.pdf
A comprehensive list of general, country-based and topic-based search engines.	www.searchenginecolossus.com/
Valuable list of both single and metasearch search engines.	www2.dynamite.com.au/kiwisunf/Search.htm

Following links

Links to numerous family and social policy sites throughout the world.	www.aifs.org.au
Links to national statistical agencies and online statistics throughout the world.	www.abs.gov.au
	Then select the **Statistics** button and then select **Links to other statistical sites**

Gateways

Social Science Information Gateway (SOSIG).	www.sosig.ac.uk/
Lists of links to social science gateways.	http://odwin.ucsd.edu
Excellent gateway to a wide range of substantive topic areas in sociology.	www.pscw.uva.nl/sociosite/topics/index.html
Extensive list of social science resources on the internet.	www.wcsu.ctstateu.edu/socialsci

---

WEB POINTER 3.2 *continued*

Subscriber sites to online journals and other publications

Annual reviews of anthropology, sociology, political science, psychology.	http://arjournals.annualreviews.org/socialhome.dtl
Access to electronic versions of a wide range of journals.	http://firstsearch.oclc.org/
Lists online journals of Sage Publications.	www.sagepub.co.uk/journals/details/c001.html
Lists online journals of publisher Taylor and Francis.	www.tandf.co.uk/journals/sublist.html
Fully online, currently free sociology journal.	www.socresonline.org.uk/

Online catalogues

Online catalogues of University Research Libraries in the UK and Ireland.	http://copac.ac.uk/
British Library.	http://blpc.bl.uk/
Canadian websites and catalogues.	www.nlc-bnc.ca/canlib/eindex.htm
European Commission Libraries Catalogues.	http://europa.eu.int/eclas/
Europe's national libraries	www.konbib.nl/gabriel/en/services.html
Yahoo! index of library sites.	http://dir.yahoo.com/Reference/Libraries/
Library web pages and online catalogues.	http://staffweb.library.vanderbilt.edu/breeding/libwebcats.html
Library servers via WWW.	http://sunsite.berkeley.edu/Libweb/
United Kingdom libraries catalogues.	www.bopac2.comp.brad.ac.uk/~bopac2/egwcgi.cgi/egwirtcl/mtargets.egw/1+0
Worldwide list of national library catalogues.	www.library.uq.edu.au/natlibs/
Web-based online catalogues.	www.dex.com

Free online articles

Access to more online newspapers.	http://library.uncg.edu/news/
Access to very large number of online newspapers.	www.onlinenewspapers.com/
Access back issues of newspapers.	www.newstrawler.com/nt/nt_home.html
Provides access to a free web clipping service that notifies re articles on nominated topics from nominated newspapers on publication.	http://nt.excite.com/

WEB POINTER 3.2 *continued*

Lists of restricted access academic journal articles

Free access to Ingenta which enables a free search of articles from a wide range of academic and other journals.	www.ingenta.com
Subscriber access to Blackwell's Electronic Journal Navigator which includes articles and abstracts for papers from a large range of journals.	http://navigator.blackwell.co.uk
Free notification service for articles published in the numerous Taylor and Francis journals.	www.tandf.co.uk/sara/

Restricted access searchable journal databases
(sites for these databases vary from library to library)

Anthropological Index of the Royal Anthropological Institute.	Free online Index to Current Periodicals in the Museum of Mankind Library
EconLit	Includes citations and abstracts from 400+ major journals plus articles in collective volumes books, dissertations, and working papers.
Medline Express	Provides bibliographic citation information and abstracts from 3700+ journals. Includes health topics relating to psychiatry and psychology, social sciences and education, technology, humanities and health care.
PsycInfo	Contains citations and summaries of journal articles, book chapters, books reports and dissertations in psychology and psychological aspects of related disciplines (1984 onwards).
Sociological Abstracts	Provides bibliographic information and abstracts to articles from journals (2500+), sociology conferences, dissertations and selected sociology books (1963 onwards).
Ebsco Megafile	A collection of full text databases covering social and health sciences.

Newsgroups

Topic-based list of newsgroups with search facility.	www.topica.com/dir/?cid=0

Listservs

List of sociology related listservs.	www.pscw.uva.nl/sociosite/mailinglists.html
Social sciences (sociology, anthropology, psychology, law, politics, economics) listservs.	www.mailbase.ac.uk/category/L.html

---

WEB POINTER 3.2   *continued*

List of listserv lists.	http://paml.alastra.com/sources.html
Extensive list of listservs.	http://paml.alastra.com/indexes.html

Visit www.social-research.org to use these links and to check for updates and additions.

SCOPE OF THE RESEARCH

As well as specifying what is to be explained and in what terms it is to be explained, it is helpful to resolve two issues related to the scope of the research before collecting data.

Particular but exhaustive or general but partial?

Will the research focus on a particular case and try to find as many causes as possible or will it seek partial explanations of a class of events? For example, a study of strikes could focus on a particular strike and try to discover all the factors behind that strike (including those peculiar to it alone). Alternatively, it could look at strikes in general and try to locate some factors that tend to be of importance for strikes in general. The first approach that focuses on obtaining thorough accounts of particular cases is called *idiographic* explanation (like idiosyncratic). Finding a partial explanation of a class of cases (e.g. strikes) is called *nomothetic* explanation. If we are committed to doing a survey then we will need to formulate the research question in terms appropriate to nomothetic explanations. If the question is really more suited to an idiographic approach we would be best advised to try another research method.

Units of analysis

A unit of analysis is the unit about which we obtain information: it is the unit whose characteristics we describe. In survey research the unit of analysis often is an individual. In a survey we might ask 2000 people whether they have ever been on strike or what might cause them to go on strike. But other units of analysis are possible. Any region (e.g. country, county, state) could be used so that we look at the strike rate of various regions for example. Time periods can be used. We might compare different years in terms of the strike levels. Events can be used

too. Thus a study could be designed around collecting data about various strikes. A group or organisation (e.g. particular unions) could be the unit of analysis. Many other units of analysis can be used depending on the issue at hand. Poems, paintings, buildings, jokes, newspapers, families and so on could be used.

Working out the unit of analysis is important in two respects. First, being aware of the range of possible units of analysis can help formulate more useful and interesting research questions and highlight a range of types of relevant data. Only to think of collecting data from and about individuals can lead to asking rather restricted research questions. Second, if data cannot be collected using a particular unit of analysis, the general thrust of the question may be retained simply by changing to a unit of analysis about which data are available. For example, we may want to know whether prosperous economic conditions encourage or discourage industrial disputes. Initially we might try to collect data from a country for the last 50 years but if this was unavailable for a particular country we could try another approach by comparing disputation levels in different countries with varying economic conditions. By changing the units of analysis from years to countries we have a different handle on the problem.

Where a number of units of analysis can be used in the one study we can be more confident in the general thrust of the results. It provides a tougher test of a theory. In a study on the effect of economic conditions on strikes it would be best to use both years and countries as units of analysis since this enables us to approach the same question from a variety of angles.

RESEARCH DESIGN

With either descriptive or explanatory research it is necessary to have a frame of reference within which

to interpret the results—a frame of reference that enables us to do more than simply report the results.

Descriptive research

When dealing with descriptive questions the need for a frame of reference is fairly obvious. For example, the inflation rate of a country might be 9 per cent. But is this high or low, good or bad, improving or deteriorating? To obtain an appreciation of this figure we need to know how it compares with the inflation rate of other countries or of other years: we need a context to make sense of most data. When collecting data it is necessary therefore to design strategies to ensure that the data needed to provide this context are collected.

Two useful types of information are data about other groups and about the same group of people over time. This information can provide a context in which to view a single piece of data and helps avoid drawing faulty conclusions. Suppose we have collected information about the levels of self-esteem of women who are out of the workforce raising young children, and that we have found that many of these women exhibit low levels of self-esteem. What does this mean? On its own it means very little. It would be much more helpful to be able to compare the findings for these women with those for some other groups. How do they compare with comparable women in the workforce? How do they compare with men who are out of the workforce raising young children? We might be tempted to conclude that being out of the workforce has a detrimental effect on women's self-esteem, but on its own the finding does not justify this conclusion. We need to know whether the self-esteem of women out of the workforce is *different* from that of other comparable groups.

As well as comparing these women with other comparable groups it would be desirable to examine their levels of self-esteem over a period of time and see what their self-esteem was like before they left the workforce, what it was like when they were out of the workforce and what it was like when they returned. Looking at the same group of people over time helps provide a context in which to understand the low self-esteem of mothers out of the workforce.

Explanatory research

When asking research questions that deal with *causal processes* an appropriate research design is absolutely essential. Where the goal is to develop an explanation of the patterns in the data we need to eliminate as many alternative explanations of the patterns as possible. If we develop a causal model that proposes that X produces Y we need to be as sure as we can that it is in fact X and not A, B or C that produces Y. Paying careful attention to research design is an important way of ensuring that possible alternative explanations of the data are eliminated. Good design also gives us more confidence in the conclusions we draw about the causal processes operating.

People use the term 'research design' in different ways. Some use it broadly to include the issues of problem formulation, operationalisation (see Chapter 4), sampling and the selection of data collection techniques. Like Stouffer (1950) I use the term much more narrowly to refer to the *structure of the data* rather than the particular data. The central point of good research design is that it provides a context in which relatively unambiguous statements can be drawn. The aim is to move our conclusions about causal processes from the realm of the plausible and possible to the convincing and compelling.

In the sections that follow I outline a number of

WEB POINTER 3.3   *Illustrations of worked through research questions*

Illustrates the refinement of several research questions and some steps in using the internet for information reviews. These are very helpful examples.

http://tramss.data-archive.ac.uk/
     ResearchQuestions/index.asp

Visit www.social-research.org to use these links and to check for updates and additions.

research designs following Stouffer's (1950) discussion. I begin with a description of the *experimental design* and then deal with a number of common variations. These and other designs are discussed in detail in my book *Research Design in Social Research* (de Vaus, 2001). Software provided in the Methodologist's Toolchest can assist in developing useful research designs for your particular project (Box 3.4). Although the experimental design is impractical for most social science research problems, it nevertheless provides a useful benchmark against which to compare other designs and highlight their weaknesses. In so doing it can help us both design and consume research more intelligently and critically. The experimental design also highlights the logic behind many of the statistical techniques covered later in this book.

BOX 3.4
Software to assist with the development of a research design

The *Methodologist's Toolchest* is a set of modules designed to assist with developing and analysing social research. One module, **Research Design**, assists in developing research designs. This module asks for details of your design and forces you to think through elements of the design. It identifies possible problems with the design and makes suggestions for improvements. This module deals with research design in a far more detailed way than is possible in this chapter. This software can be purchased from: www.scolari.co.uk/toolchest/tootlchest.htm.

The classic experimental design

In its simplest form the experimental design has two groups: an experimental group and a control group. It also extends over time so that data are collected at, at least, two points in time (before and after). Between Time 1 (before) and Time 2 (after) the experimental group is exposed to an experimental intervention. The control group is left alone. At both Time 1 and Time 2 the experimental and control groups are measured in relation to the key dependent variable that is of interest in the study. In Figure 3.2 the measure of the dependent variable is indicated by E_1 and E_2, and C_1 and C_2.

Since the experimental group has been exposed to the experimental intervention, we might suppose that differences in the experimental group between Time 1 and Time 2 might be due to the influence of the intervention. To test for an intervention effect we measure the difference between E_1 and E_2 and calculate the change between Time 1 and Time 2 (E_{Change}) for the experimental group.

However, the observed change for the experimental group between Time 1 and Time 2 might be due to factors other than the experimental intervention. A change could occur due to the passing of time, be a result of being measured at Time 1, or be caused by a whole set of other possibilities. For this reason a control, or comparison, group is needed. Ideally this group should be identical to the experimental group at Time 1. However, unlike the experimental group, they are not exposed to any experimental intervention. We can measure them on our dependent variable at both Time 1 and Time 2 and obtain a measure of change over that time (C_{Change}). Since this group was not exposed to the experimental intervention, any change in this group will not be due to this factor.

The crucial thing to look at is *whether the experimental group changed more than the control group*. If the

Method of allocation to groups	Time 1 (T_1) Before	Intervention (X)	Time 2 (T_2) After	
Experimental group (Random allocation)	E_1	Intervention	E_2	$E_{Change} = E_2 - E_1$
Control group (Random allocation)	C_1	No intervention	C_2	$C_{Change} = C_2 - C_1$

Effect = $E_{Change} - C_{Change}$

Figure 3.2   Structure of the classic experimental design

experimental group changed significantly more, we normally would conclude that this is because of the experimental intervention.

Of course, this conclusion is warranted only if both groups were the same to start with and had identical experiences between Time 1 and Time 2. In order to ensure that they are the same to start with, people are *randomly assigned* to the experimental and control groups. So long as the groups are large enough, random assignment should ensure that they are very similar in most respects. Ensuring that they have the same experiences between Time 1 and Time 2 is more difficult and is one reason why experiments are often held in laboratories where this can be controlled to some extent.

An example of an experimental design is illustrated in Figure 3.3. Here the research is designed to test whether an intervention to help smokers stop smoking is effective. We would begin with a pool of people who smoked. We would then randomly allocate these smokers to either an experimental group (those who will do the QUIT smoking program designed to help smokers stop smoking) and the control group (will not do the program). This random allocation should ensure that the two groups would initially smoke to a similar extent and be alike in other ways (age, duration of smoking, gender etc). We would measure the level of smoking in each group before the QUIT program began. Six months after the experimental group had completed their program we would remeasure the percentage in each group who were still smoking. In this example we would observe that in the group that took the program 10 per cent fewer smoke by Time 2. In the control group 3 per cent fewer smoke at Time 2 than at Time 1.

Since there was a reduction of 3 per cent among the control group it is likely that the experimental group would have seen a similar decline in smoking in the same period without the QUIT program. In other words, some of the 10 per cent drop in smoking among the experimental group is likely to be due to factors separate from the program. The effect of the program would be measured by the *difference* in the amount of change between the experimental and control groups (i.e. the program was responsible for a 7 per cent decline in smoking).

By designing the research with the before and after dimension and an experimental and control group, we are in a position to draw much more unequivocal conclusions than we could without that design. We are much more able to draw conclusions about causal processes.

There are a number of problems which make it difficult to use an experimental design for social research. In many situations it is not possible to obtain repeated measures for the same group, thus making it impossible to get measures at both Times 1 and 2. Often it is difficult to obtain a control group. Practical and ethical considerations often make it impossible to introduce experimental interventions. For example, we might be interested in the effect of marital breakdown on the social adjustment of young children. Obviously we cannot allocate people randomly to two groups and somehow cause marital breakdowns in one group and leave the other group alone.

Panel design

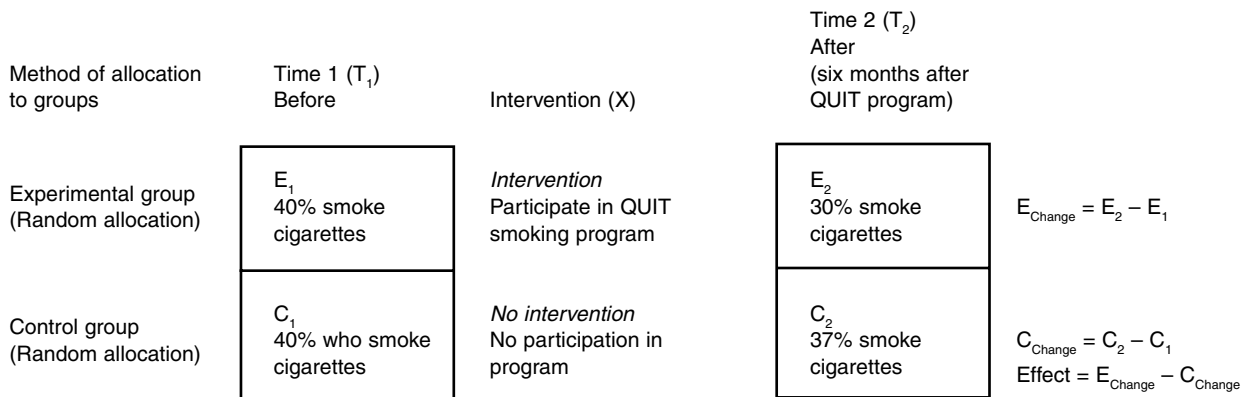This design (see Figure 3.4) uses only the top two cells of the experimental design. It looks at the same

Method of allocation to groups	Time 1 (T_1) Before	Intervention (X)	Time 2 (T_2) After (six months after QUIT program)	
Experimental group (Random allocation)	E_1 40% smoke cigarettes	*Intervention* Participate in QUIT smoking program	E_2 30% smoke cigarettes	$E_{Change} = E_2 - E_1$
Control group (Random allocation)	C_1 40% who smoke cigarettes	*No intervention* No participation in program	C_2 37% smoke cigarettes	$C_{Change} = C_2 - C_1$ Effect $= E_{Change} - C_{Change}$

Figure 3.3   Experimental design to ascertain the effect of a QUIT program on smoking
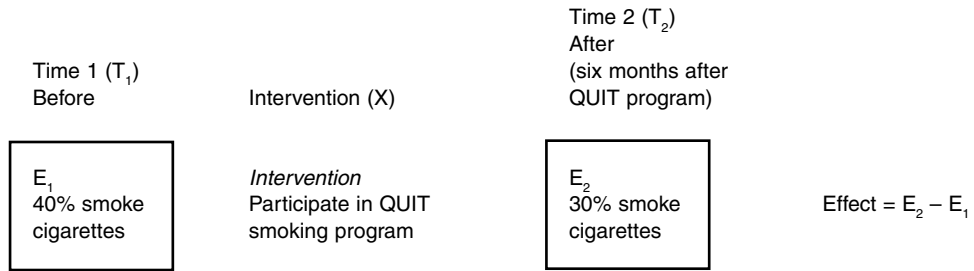
Figure 3.4   Panel design to ascertain the effect of a QUIT program on smoking

group of people over a period of time: we would measure smoking levels at both points and the whole group would participate in the QUIT program. The difference in smoking levels between Time 1 and Time 2 provides a measure change over the period.

We might be tempted to conclude that this change is due to the intervention—the QUIT program. However this design does not allow us to draw this conclusion. We need to know the extent to which comparable smokers who did not participate in the QUIT program stopped smoking. There may well have been other events occurring between Time 1 and Time 2 (e.g. increase in price of cigarettes, increased unemployment, new brands released, new marketing strategies of the tobacco companies) that might be responsible for the observed changes in smoking levels.

Quasi-panel design

This is similar to the panel design except that *different* groups of people are studied at the two points of time (see Figure 3.5). This design might be used to avoid the difficulties of keeping track of the same people over time. Using the above example, the

quasi-panel design would involve measuring smoking behaviour of a representative sample of people before implementing a widespread anti-smoking campaign which includes easy access to QUIT programs. After the campaign we would remeasure the smoking behaviour of another representative sample. Any decline in smoking behaviour evident between the two samples might be attributed to the anti-smoking campaign.

However, this design has the same problems of the panel design for drawing causal inferences. It also has the additional problem of being unable to fully match the samples at Time 1 and Time 2. Differences observed between Time 1 and Time 2 might be due to sampling error—differences between the samples. We cannot even be sure that we have measured change—something that the panel design is at least good at.

Retrospective panel design

The panel design has the disadvantage of having to keep track of people over time and of having to wait for a long time on occasions before the results can be collected. Often it is simply not feasible to follow a
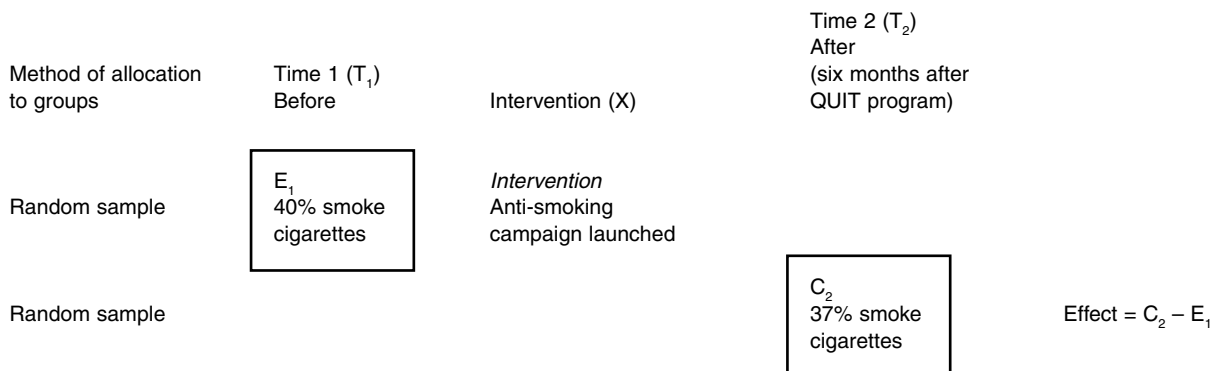

Figure 3.5   Quasi-panel design to ascertain the effect of an anti-smoking campaign on smoking

group of people over time and the drop–out rate in the study can create serious difficulties with comparing Time 1 and Time 2 results.

An alternative approach that can be used to provide a time dimension to the study is to obtain information at one time only but to ask about two or more time points and to find out about events in between these two times (see Figure 3.6). Thus we might ask a sample of people about their current level of smoking and about their level six months earlier (i.e. before an anti-smoking campaign was launched).

Apart from the other shortcomings of the panel design, this design has the difficulty of selective memory and is open to the possibility that people will reinterpret the past in the light of the present. People can mis-remember the past and become confused about time periods. What they remember as six months ago might be two years ago (telescoping) or only two months ago (reverse telescoping) (Menard, 1991). Even if such a study detects 'changes'

we cannot be sure whether the changes are real or perceived.

Retrospective experimental design

The retrospective panel design suffers from the dual problem of the shortcomings of retrospectivity and from having no control group. The first problem means that we cannot be certain about the extent of real change while the absence of a control group means that it is difficult to draw conclusions about causal processes. The retrospective experimental design (see Figure 3.7) is an attempt to deal with the control group problem.

Using this approach we would interview a sample of people and ascertain their level of awareness of the recent anti-smoking campaign (or whether they participated in a QUIT program). We would then ask about their current level of smoking and about their level six months earlier before the campaign. We would then extract from our sample
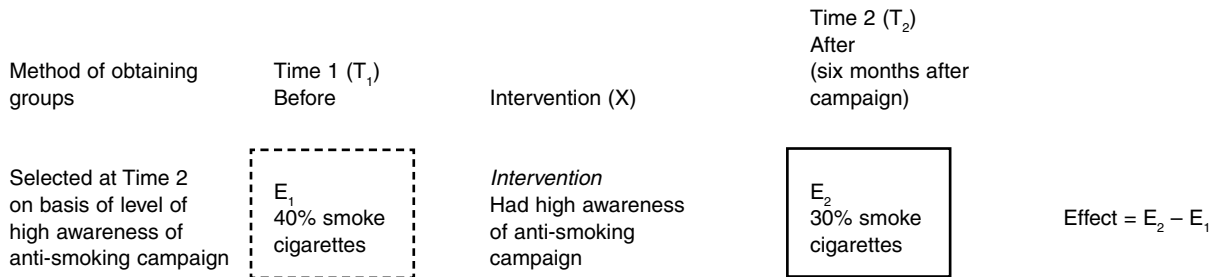
Figure 3.6   Retrospective panel design to ascertain the effect of an anti-smoking campaign on smoking
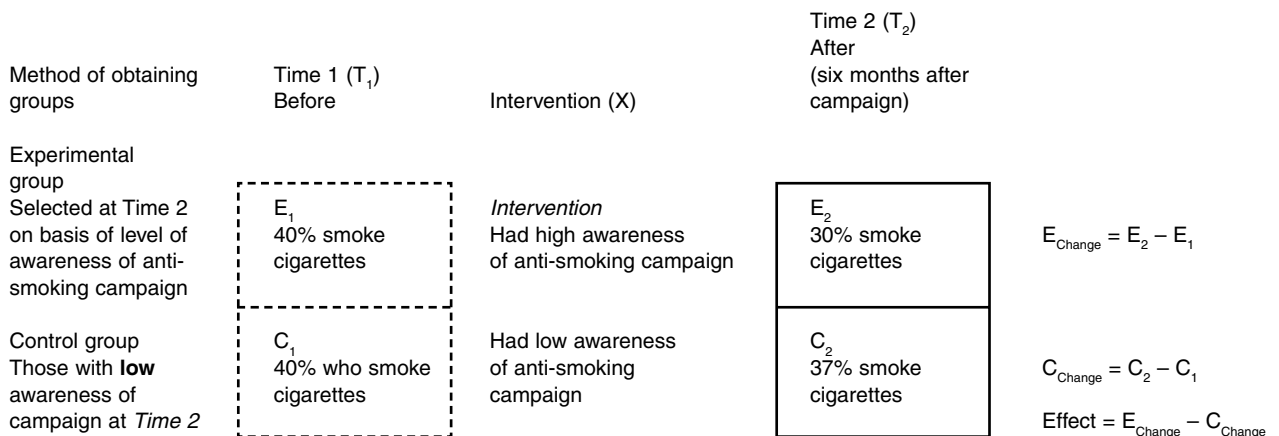
Figure 3.7   Retrospective experimental design to ascertain the effect of an anti-smoking campaign on smoking

those who had a high awareness of the campaign and then extract a group with low awareness. We would try to ensure that the low awareness group were matched in other regards (age, sex, class etc) with the high awareness group. If the level of smoking in the high awareness group declined more than in the low awareness group we might attribute this to the effect of the campaign.

Cross-sectional or correlation design

The most common design used in survey research is the cross-sectional design (see Figure 3.8). Using this approach we collect measures from at least two groups of people at one point of time and compare the extent to which the two groups differ on the dependent variable (e.g. levels of smoking).

For our example we would obtain a random sample of people after an anti-smoking campaign had ended. We would ask these people about their current level of smoking and how aware they were of the anti-smoking campaign. We could then divide people into groups according to how well aware they had been of the campaign. If the campaign was effective we would expect that those with the greatest awareness would also have the lowest levels of smoking. If this turned out to be the case we may be tempted to conclude that the campaign was effective—that it produced a lower level of smoking.

The problem is that the high and low awareness people might differ in other ways apart from their awareness of the campaign. High awareness people might be older, be in poorer health or be distinctive from the low awareness group in some way. *These differences, rather than high awareness, might have* produced the lower level of smoking. Alternatively, non-smokers may be more conscious than smokers of anti-smoking campaigns. This, rather than the effectiveness of the campaign, would explain the different levels of smoking of the high and low awareness groups. We would need to eliminate these possibilities before we could begin to draw conclusions about the effectiveness of the campaign. Methods of achieving this are discussed in Chapters 16 and 17.

Although the discussion of research designs has been limited to two groups this is not a requirement. Any number of groups can be built into a design. In cross-sectional designs this is particularly common. If we were interested in the effect of campaign awareness on smoking we could have two groups— high and low awareness—and see whether their smoking differed. But we could have, say, six awareness groups ranging from high to low. To the extent that the six groups differ, we might conclude, other things being equal, that campaign awareness is related to smoking levels.

One group post-test only design

The most primitive design—of little use when trying to analyse causal processes—is one that has only one cell of the experimental design (see Figure 3.9). This design involves collecting information from one group at one point in time. For our study of smoking behaviour it would involve asking a sample about their level of smoking some time after an anti-smoking campaign. We might observe that only 30 per cent smoke. Can we tell anything about the effectiveness of the campaign? Not a thing. To use this design for causal analysis requires that we rely on
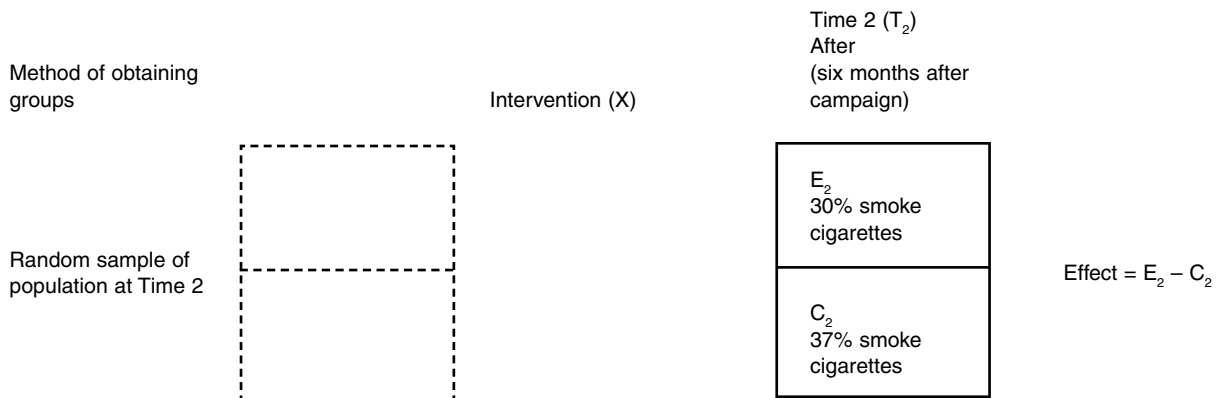

Method of obtaining groups		Intervention (X)	Time 2 (T_2) After (six months after campaign)	
			E_2 30% smoke cigarettes	
Random sample of population at Time 2				Effect = $E_2 - C_2$
			C_2 37% smoke cigarettes	

Figure 3.8 Cross-sectional design to ascertain the effect of an anti-smoking campaign on smoking

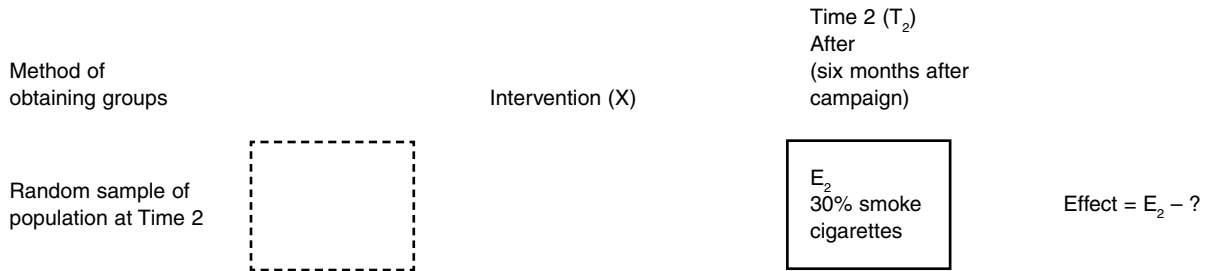| Method of obtaining groups	Intervention (X)	Time 2 (T_2) After (six months after campaign)	
Random sample of population at Time 2	[dashed box]	E_2 30% smoke cigarettes	Effect = E_2 – ?

Figure 3.9 **One group post-test only design to ascertain the effect of an anti-smoking campaign on smoking**

plausible conjecture and our preconceptions and assumptions about the data that *might* be in the 'missing' cells of the design. It requires that we fill in the cells from our imagination rather than from systematic observation. Without an empirical frame of reference against which to compare the 30 per cent figure we can not say anything about causal processes.

Why bother about design?

The aim of this outline of research design has been fourfold. First, rather than suggesting that the experimental design is the only way to proceed, the intention has been to create an awareness of the limitations of designs where cells are missing and to suggest ways of filling in at least some of the cells. The more cells that are filled in with data the better the design will be. In many situations in social research the experimental design will either be impracticable or impossible. Second, the aim has also been to highlight the logic behind the experimental design, for it is in this logic that the rationale of multivariate analysis lies. The use of multivariate analysis with a cross-sectional design is the most common way in which survey researchers attempt to approximate the logic of the classical experiment. Understanding design principles therefore lies at the heart of an understanding of the more sophisticated methods of statistical analysis. Third, an understanding of the principles of research design should help us to be more critical consumers of social research. Finally, an appreciation of design issues highlights the importance of what data to collect (e.g. do we need to ask retrospective questions?) and of thinking through the issues of whom we should collect it from. It is therefore important in questionnaire design and in developing sampling strategies.

BOX 3.5
Software to assist the development of research proposals

The *Methodologist's Toolchest* includes a very helpful module called **The Peer Review Evaluator** which assists in developing research proposals. Among other things, it takes you through the many elements of well-defined research problems and provides invaluable assistance in taking care of the many processes and points out areas you have not dealt with. This software is available for purchase at: www.scolari.co.uk/toolchest/tootlchest.htm. The Peer Review Evaluator assists with addressing many of the questions below and many more.

REFINING RESEARCH QUESTIONS: A CHECKLIST

To help clarify precisely what your research question is, try to be clear about your answers to the following groups of questions:

1 Aims
  a What type of study are you doing: Is it descriptive or explanatory?
  b What will be the main type of contribution of the study? Is it intended to make a practical, theoretical/conceptual or methodological advance?
  c Is the study testing well-formulated hypotheses or simply asking a set of questions? What are they?

d How will the study add to what is already known? Are you:

   i testing a new theory, hypotheses, methods?

   ii improving on previous methods and theories?

   iii testing alternative theories?

   iv trying to resolve conflicting evidence from previous research?

   v measuring concepts in a new way?

   vi applying the ideas in a different context and to different types of people and groups than previously?

   vii seeking to replicate or challenge previous research?

e Are the ideas you are testing reasonable? Do they make sense in terms of previous research, theory and experience?

f Are the questions answerable or testable? Is it possible to measure the core concepts (see Chapter 4)? Do you have access to a sample required for doing the study?

2 Context

a How does the study relate to previous research on the topic? Is it a replication, a test, an extension? What is different about it?

b How adequate is the previous research? What are the strengths and weaknesses of previous studies? Is there much previous research? Is the previous research consistent or are there conflicting findings and contradictions to be resolved?

c Have you critically evaluated the previous research? Your review of the research should not simply list what has been done. It should point out the gaps, limitations and other shortcomings. Does your research question emerge out of the shortcomings and gaps (empirical, theoretical or methodological) you have identified in the previous research? If it does not why are you doing the research?

d Have you examined the most recent research? How up to date is your review? Have you looked at books, reports, work in progress as well as journal articles?

e How balanced is your review of previous research? Have you simply reported research that confirms your views or have you reported fairly on the full range of available findings?

3 Significance and scope

a Why is the study worth doing? Of what importance will the results be? If you plan to make a practical, theoretical or methodological contribution of what significance is this likely to be?

b What are the theories you will be evaluating? What concepts will you be using? Are these important?

c If you are evaluating theories are you clear what these theories are? Have you drawn a diagram of the theory?

d What is the scope of the study? Will the findings be widely generalisable?

e To which wider population or subgroups will the results apply?

f What will be your unit of analysis? Can you add to previous research by using different units of analysis?

g To what time period will the results apply?

h To what region(s) will the results apply?

4 Research design

a If you are conducting an explanatory study have you specified the type of research design you will be using?

b What comparison groups will you be using?

c Will your sample have sufficient diversity so that you can build comparison groups for data analysis? How will you make groups comparable?

d If you are looking at change do you have a time dimension to your study? Is the time over which you will look at change sufficient? Will you be using a prospective or retrospective approach?

KEY CONCEPTS

Cross–sectional design	Idiographic	Nomothetic	Research design
Dependent variable	Independent variable	Panel design	Unit of analysis
Experimental design	Intervening variable	Quasi–panel design	Variable
Explanation	Intervention	Random allocation	

FURTHER READING

Remarkably few books seriously address how to clarify research problems. The books by Bell and Newby, *Doing Sociological Research* (1977) and Bell and Encel, *Inside the Whale* (1978) are valuable because they show that in actual practice the formulation of research problems is a process.

Babbie, in *The Practice of Social Research* (1995), provides a useful account of options and decisions when designing a research project. Locke et al. *Guide for Planning Dissertations and Grant Proposals* (1999) provides practical advice about focusing research questions. Cooper's book, *Synthesising Research: A Guide for Literature Reviews* (1998), and his earlier article on 'Literature Searching Strategies of Integrative Research Reviewers' (1985) and Fink's book *Conducting Research Literature Reviews: From Paper to the Internet* (1998) provide some useful hints on reviewing the literature as a way of clarifying research questions. Gallagher's recent article on the significant role of the literature review in designing research is also worth looking at.

An advanced book outlining a variety of research designs that is well worth the effort of reading is Campbell and Stanley's *Experimental and Quasi-Experimental Designs for Research* (1963). However it is worth reading Stouffer's paper on 'Some observations on study design' (1950) first. *Research Design in Social Research* (de Vaus, 2001) provides an extensive treatment of a wide range of experimental, panel and cross-sectional designs and examines the strengths and weaknesses of each and the way each design affects the way data are analysed.

EXERCISES

1   For each of the following statements say what unit of analysis is being used.

   a   In the UK for every 1000 women aged 20–24 there were 30.4 who had an abortion in that year of 1998.

   b   In 1998 in the United States the average family in poverty would require an additional US$6620 per year to get on or above the poverty line.

   c   Australia has one of the lowest rates of expenditure on research amongst developed countries.

   d   Within any one year 18 per cent of Australians move.

   e   In the UK the official abortion rate per 1000 women aged 20–24 has changed as follows:

1968 = 3.4	1985 = 20.4
1970 = 10.5	1990 = 28.1
1975 = 15.1	1995 = 25.5
1980 = 18.7	1998 = 30.4

2   You wish to see whether having a secure family income actually makes families function better. Forgetting for the moment how you would measure 'better functioning' or 'secure income', draw design diagrams of a cross–sectional and a panel design to test this idea. Then draw a design that avoids the problems inherent in these two.

3   You wish to do a descriptive study on prejudice. Using the guidelines suggested in Chapter 3, focus this topic into a much more specific research problem.

4   Try putting the theory which is diagrammed in Figure 3.1 into words.

5   For each of the following statements of research findings indicate the type of research design that appears to have been employed and explain what is wrong with the conclusions that are drawn. Concentrate on problems that arise from research design problems.

   a   Sixty-eight per cent of married people scored high on our index of conservatism while only 38 per cent of single people scored high. Marriage makes people more conservative.

b   After observing a sample of childless married couples over a ten-year period we observed that the level of marital happiness declined over this period. Childlessness works against people being happily married.

c   In the early 1970s, before the end of the Vietnam War, surveys showed that tertiary students had strong anti–American attitudes. Recent surveys have shown that these feel-ings are no longer evident among students. Ending the Vietnam War certainly improved the attitudes of students to the United States.

d   Old people attend church more often than young people. For example, 58 per cent of those over 60 attend church regularly while only 22 per cent of those under 25 do so. From this we can conclude that as people get older they become more religious.

e   The average number of children per family now is 1.8. Families are obviously getting smaller these days.

f   To test the idea that having children makes people happier, a group of parents were asked how happy they felt now compared with before they had children. Eighty-seven per cent said they were happier now than before they had children. From this we can conclude that having children improves people's happiness.

g   A HEADSTART program (a preschool educational program to help disadvantaged children have a head start by the time they commence school) was used to test the effectiveness of HEADSTART. A group of four-year-olds from disadvantaged back-grounds were chosen to enter the program. IQ tests were given at the beginning of the program and again at the end. There was an average gain of ten IQ points over the period of the program. HEADSTART increases children's IQ.