PART THREE

IMPLICATIONS

7

Validity and Reliability

Much of the textbook debate about the scientific status of sociology is somewhat fatuous. Few researchers would now dispute that the cultural world has different properties from the natural world. Again, only hard-core laboratory scientists would assume that the controlled experiment offers an appropriate or indeed useful model for social science.

It is an increasingly accepted view that work becomes scientific by adopting methods of study *appropriate* to its subject matter. Sociology is thus scientific to the extent that it uses appropriate methods and is rigorous, critical and objective in its handling of data. As Kirk and Miller argue:

> The assumptions underlying the search for objectivity are simple. There is a world of empirical reality out there. The way we perceive and understand that world is largely up to us, but the world does not tolerate all understandings of it equally. (Kirk and Miller: 1986, 11)

Kirk and Miller remind us of the need for 'objectivity' in scientific research. It is particularly important to remember this need in social research. The array of suggestive theories and contrasting methodologies, reviewed in Part Two of this book, may tempt us to believe that 'anything goes'. However, such anarchy is, I believe, bad for social research in at least two ways. First, it leads us to believe that the only important debates are conducted between 'armchair' theorists. Second, by downplaying the cumulative weight of evidence from social science research, it lowers our standing in the community.

The real issue is how our research can be *both* intellectually challenging and rigorous and critical. One way of being critical is, as Popper (1959) has suggested, to seek to refute assumed relations between phenomena. This means overcoming the temptation to jump to easy conclusions just because there is some evidence that seems to lead in an interesting direction. Instead, we must subject this evidence to every possible test. Then, only if we cannot refute the existence of a certain relationship, are we in a position to speak about 'objective' knowledge. Even then, however, our knowledge

is always provisional, subject to a subsequent study which may come up with disconfirming evidence.

Popper puts it this way:

> What characterises the empirical method is its manner of exposing to falsification, in every conceivable way, the system to be tested. Its aim is not to save the lives of untenable systems but, on the contrary, to select the one which is by comparison the fittest, by exposing them all to the fiercest struggle for survival. (Popper: 1959, 42)

The two central concepts in any discussion of rigour in scientific research are 'reliability' and 'validity'. I will discuss each in turn, examining what each concept means in practice in both quantitative and field research.

Reliability

> [Reliability] refers to the degree of consistency with which instances are assigned to the same category by different observers or by the same observer on different occasions. (Hammersley: 1992a, 67)

What reliability involves and its relation to validity can be understood simply following Kirk and Miller's example of using a thermometer:

> A thermometer that shows the same reading of 82 degrees each time it is plunged into boiling water gives a reliable measurement. A second thermometer might give readings over a series of measurements that vary from around 100 degrees. The second thermometer would be unreliable but relatively valid, whereas the first would be invalid but perfectly reliable. (1986, 19)

Kirk and Miller usefully distinguish three kinds of reliability, as follows:

1 *Quixotic reliability*: 'the circumstances in which a single method of observation continually yields an unvarying measurement'; but this kind of reliability can be 'trivial and misleading'. For instance, just because an interview question always elicits a predictable response does not mean that it is analytically interesting or that the response necessarily relates to what people say and do in different contexts.

2 *Diachronic reliability*: 'the stability of an observation through time'. For instance, showing that ways of defining advice sequences work equally well with data drawn from different periods.

3 *Synchronic reliability*: 'the similarity of observations within the same time-period' (Kirk and Miller: 1986, 41–42). A standard way through which this is assessed is through triangulation of methods (e.g. the use of interviews as well as observation). As Kirk and Miller argue, paradoxically the value of such triangulation is that it 'forces the ethnographer to imagine how multiple, but somehow different, qualitative measures might simultaneously be true' (1986, 42).

Not a Problem?

Some social researchers argue that a concern for the reliability of observations arises only within the quantitative research tradition. Because such 'positivist' work sees no difference between the natural and social worlds, it is appropriately concerned to produce reliable measures of social life. Conversely, it is argued, once we treat social reality as always in flux, then it makes no sense to worry about whether our research instruments measure accurately.

This is an example of such an argument:

> Positivist notions of reliability assume an underlying universe where inquiry could, quite logically, be replicated. This assumption of an unchanging social world is in direct contrast to the qualitative/interpretative assumption that the social world is always changing and the concept of replication is itself problematic. (Marshall and Rossman: 1989, 147)

But is this so? It is one thing to argue that the world is processual; it is much more problematic to imply, as Marshall and Rossman seem to do, that the world is in infinite flux (appropriate to the pre-Socratic philosopher Heraclitus, perhaps, but not a comfortable position for social scientists).

Such a position would rule out any systematic research since it implies that we cannot assume any stable properties in the social world. However, if we concede the possible existence of such properties, why shouldn't other work replicate these properties? As Kirk and Miller argue:

> Qualitative researchers can no longer afford to beg the issue of reliability. While the forte of field research will always lie in its capability to sort out the validity of propositions, its results will (reasonably) go ignored minus attention to reliability. For reliability to be calculated, it is incumbent on the scientific investigator to document his or her procedure. (Kirk and Miller: 1986, 72)

Following Kirk and Miller, I consider below how reliability can be addressed in qualitative studies. I will look in turn at the four methodologies discussed in Part Two of this book: observation, textual analysis, the interview and the transcript of naturally-occurring talk.

Reliability and Observation

Observational studies rarely provide readers with anything other than brief, persuasive, data extracts. As Bryman (1988) notes about the typical ethnography: 'field notes or extended transcripts are rarely available; these would be very helpful in order to allow the reader to formulate his or her own hunches about the perspective of the people who have been studied' (Bryman: 1988, 77).

Although, as Bryman suggests, extended extracts from fieldnotes would be helpful, the reader also should require information on how fieldnotes were recorded and in what contexts. As Kirk and Miller argue: 'The contemporary search for reliability in qualitative observation revolves around detailing the relevant context of observation' (Kirk and Miller: 1986, 52).

Spradley (1979) suggests that observers keep four separate sets of notes:

1 Short notes made at the time.
2 Expanded notes made as soon as possible after each field session.

3 A fieldwork journal to record problems and ideas that arise during each stage of fieldwork.
4 A provisional running record of analysis and interpretation (discussed by Kirk and Miller: 1986, 53).

Spradley's suggestions help to systematise fieldnotes and thus improve their reliability. Implicit in them is the need to distinguish between *etic* analysis (based on the researcher's concepts) and *emic* analysis (deriving from the conceptual framework of those being studied). Such a distinction is employed in the set of fieldnote conventions set out in Table 7.1.

Table 7.1: *Some Fieldnote Conventions*

Sign	Convention	Use
" "	double quotation marks	verbatim quotes
' '	single quotation marks	paraphrases
( )	parentheses	contextual data or fieldworker's interpretations
⟨ ⟩	angled brackets	*emic* concepts
/	slash	*etic* concepts
———	solid line	partitions time

Source: adapted from Kirk and Miller: 1986, 57

Exercise 7.1

This exercise asks you to use the fieldnote conventions set out in Table 7.1. You should gather observational data in any setting with which you are familiar and in which it is relatively easy to find a place to make notes (you may return to the setting you used for Exercise 3.2). Observe for about an hour. Ideally, you should carry out your observations with someone else who also is using the same conventions.

1 Record your notes using these fieldnote conventions. Compare your notes with your colleague's. Identify and explain any differences.
2 What conventions were difficult to use? Why was this so (e.g. because they are unclear or inappropriate to the setting)?
3 Can you think of other conventions that would improve the reliability of your fieldnotes?
4 What have you gained (or lost) compared to earlier observational exercises (e.g. Exercise 3.2)?
5 Which further fields of enquiry do your fieldnotes suggest?

Reliability and Texts

When you are dealing with a text, the data are already available, unfiltered through the researcher's fieldnotes. Issues of reliability now arise only through the *categories* you use to analyse each text. It is important that these categories should be used in a *standardised* way, so that any researcher would categorise in the same way.

A standard method of doing this is known as 'inter-rater reliability'. It involves giving the same data to a number of analysts (or raters) and asking them to analyse it according to an agreed set of categories. Their reports are then examined and any differences discussed and ironed out.

In order to see how this method works, you should find a colleague who worked on the same exercise in Chapter 4. Compare your analyses of the same data and see if you can iron out any differences.

Reliability and Interviews

The reliability of interview schedules is a central question in quantitative methods textbooks. According to these books, it is very important that each respondent understands the questions in the same way and that answers can be coded without the possibility of uncertainty. This is achieved through a number of means, including:

- thorough pre-testing of interview schedules
- thorough training of interviewers
- as much use as possible of fixed-choice answers
- inter-rater reliability checks on the coding of answers to open-ended questions.

Exercise 7.2

This exercise gives you the opportunity to assess the reliability of your analysis of the data used in earlier exercises, using the method of inter-rater agreement.

You should find a colleague who carried out the same data-analysis exercises in Chapters 4–6. Return to your answers to one of those exercises and now consider:

1 What are the major differences and similarities in the way in which you used concepts and categories in this exercise?
2 Which part of either person's analysis needs to be revised or abandoned?
3 Do similarities in your analyses mean that the concepts and categories you have used are good ones (distinguish issues of reliability and usefulness)?
4 Do any differences mean that the concepts and categories you have used are badly designed and/or that you have used them inappropriately?
5 What have you learned from this comparison? How would you redo your analysis in the light of it?

In Chapter 5, I argued that a concentration on such matters tended to deflect attention away from the theoretical assumptions underlying the meaning that we attach to interviewees' answers. Nonetheless, this does not mean that we can altogether ignore conventional issues of reliability, even if we deliberately avoid treating interview accounts as simple 'reports' on reality. For instance, even when our analytic concern is with narrative structure or membership categorisation, it is still helpful to pre-test an interview schedule and to compare how at least two researchers analyse the same data.

Reliability and Transcripts

Kirk and Miller's suggestion that the conventionalisation of methods for recording fieldnotes offers a useful method for addressing the issue of reliability can be applied to transcripts. For we need only depend upon fieldnotes in the absence of audio- or video-recordings. The availability of transcripts of such recordings, using standard conventions, satisfies Kirk and Miller's proper demand for the documentation of procedures.

In conversation analysis, as discussed in Chapter 6, a method similar to inter-rater comparison is used to ensure reliability. Wherever possible, group data-analysis sessions are held to listen to (or watch) audio- or video-recordings. It is important here that we do not delude ourselves into seeking a 'perfect' transcript. Transcripts can always be improved and the search for perfection is illusory and time-consuming. Rather the aim is to arrive at an agreed transcript, adequate for the task at hand. A further benefit arising from such group sessions is that they usually lead to suggestions about promising lines of analysis.

Validity

> By validity, I mean truth: interpreted as the extent to which an account accurately represents the social phenomena to which it refers. (Hammersley: 1990, 57)

Proposing a purportedly 'true' statement involves the possibility of two kinds of *error* which have been clearly defined by Kirk and Miller (1986, 29–30):

Type 1 error is believing a statement to be true when it is not (in statistical terms, this means rejecting the 'null hypothesis', i.e. the hypothesis that there is no relation between the variables).
Type 2 error is rejecting a statement which, in fact, is true (i.e. incorrectly supporting the 'null hypothesis').

Validity in Quantitative Research

In quantitative research, a common form of Type 1 error arises if we accept a 'spurious' correlation. For instance, just because X seems always to be followed by Y, this does not mean that X necessarily *causes* Y. There might be a third factor, Z, which produces both X and Y. Alternatively, Z might be an 'intervening variable' which is caused by X and then influences Y (see Selltiz *et al*: 1964, 424–431).

The quantitative researcher, however, can use sophisticated means to guard against the possibility of spurious correlations. For instance, Lipset *et al* (1962), were aware that a correlation between membership of a printers' club and stated activity in union elections might be spurious. Perhaps people who joined such clubs were more interested anyway in union politics and so already predisposed to participate more in elections than non-members. Consequently, Lipset compared the participation rates of members and non-members who had the *same* prior interest in union politics. The results are shown in Table 7.2.

Table 7.2: *Club Membership and Voting in Union Elections*

	Political interest		
	High	Medium	Low
Club member	63%	41%	41%
Non-member	52%	28%	19%
	Percentage participating in elections		

Source: adapted from Lipset *et al*: 1962

Table 7.2 should be read vertically. It shows that, if you only compare people with the *same* political interest, club membership is associated with a larger percentage who participate in union elections. Consequently, the researchers had excluded one variable which might have rendered spurious the correlation between club membership and voting.

Following Popper's emphasis on attempts at refutation, Lipset *et al* could be reasonably confident that they had found a nonspurious correlation. However, since other factors could still be relevant in this correlation, as in any statistical finding, they could only talk about an *association* between phenomena not a *causal* relationship.

Lipset's attempt to control for spurious correlations was possible because of the quantitative style of his research. This had the disadvantage of being dependent upon survey methods with all their attendant difficulties. As Fielding and Fielding argue: 'the most advanced survey procedures themselves only manipulate data that had to be gained at some point by asking people' (Fielding and Fielding: 1986, 12).

As we saw in Chapter 5, what people say in answer to interview questions does not have a stable relationship to how they behave in naturally-occurring situations. Again, Fielding and Fielding make the relevant point: 'researchers who generalise from a sample survey to a larger population ignore the possible disparity between the discourse of actors about some topical issue and the way they respond to questions in a formal context' (*ibid*, 21).

Exercise 7.3

Table 7.3 is also drawn from the Lipset study. It relates voting in union politics to having friends who are also printers. Examine it carefully and then answer the questions beneath it.

Table 7.3

	Political interest		
	High	Medium	Low
Printer friends	61%	42%	26%
No printer friends	48%	22%	23%
	Percentage participating in elections		

Source: adapted from Lipset *et al*: 1962

1. Does Table 7.3 show that there is an association between having a printing friend and participating in union elections? Explain carefully, referring to the table.
2. Can we be confident that the degree of political interest of a printer does not make any correlation between friendships and participation into a spurious one?
3. Compare Table 7.3 with Table 7.2. Among what levels of political interest is any association between voting and (a) printer friends and (b) club membership most marked?
4. What might explain the differences between the two tables in the groups with shared political interest who are most influenced by variables (a) and (b) above?
5. How might you test this explanation?

Let me summarise what I have been saying so far. First, the criterion of *refutability* is an excellent way to test the validity of any research finding. Second, *quantitative* researchers have a sophisticated armoury of weapons to assess the validity of the correlations which they generate. Third, we should not assume that techniques used in quantitative research are the *only* way of establishing the validity of findings from qualitative or field research.

This third point means that a number of practices which originate from quantitative studies may be *inappropriate* to field research. The following assumptions are highly dubious in qualitative research:

1. All social science research can only be valid if based on experimental data, official statistics or the random sampling of populations.
2. Quantified data are the only valid or generalisable social facts.

3 Having a cumulative view of data drawn from different contexts allows us, as in trigonometry, to triangulate the 'true' state of affairs by examining where the different data intersect.

Each of these assumptions has a number of defects, many of which are discussed (and some displayed) in a number of texts concerned with qualitative research methodology, from Cicourel (1964) through Denzin (1970) to Schwartz and Jacobs (1979), Hammersley and Atkinson (1983) and Gubrium (1988).

Following the same order as in the list above, I note that:

1 Experiments, official statistics and survey data may simply be inappropriate to some of the tasks of social science. For instance, they exclude the observation of 'naturally-occurring' data by ethnographic case-studies (see Chapter 3) or by conversation and discourse analysis (see Chapter 6).
2 While quantification may *sometimes* be useful, it can both conceal as well as reveal basic social processes. Consider the problem of counting attitudes in surveys. Do we all have coherent attitudes on any topics which await the researcher's questions? And how do 'attitudes' relate to what we actually do – our practices? Or think of official statistics on cause of death compared to studies of the officially organised 'death work' of nurses and orderlies (Sudnow: 1968a) and of pathologists (Prior: 1987). Note that this is *not* to argue that such statistics may be biased. Instead, it is to suggest that there are areas of social reality which such statistics cannot measure.
3 Triangulation of data seeks to overcome the context-boundedness of our materials at the cost of analysing their sense in context. For purposes of social research, it may simply not be useful to conceive of an over-arching reality to which data, gathered in different contexts, approximates.

So my support for critical field research which takes seriously issues of validity is not based on an uncritical acceptance of the standard recipes of conventional methodology texts or the standard practices of purely quantitative research. In any event, quantitative measures offer no simple solution to the question of validity:

> ultimately all methods of data collection are analysed 'qualitatively', in so far as the act of analysis is an interpretation, and therefore of necessity a selective rendering. Whether the data collected are quantifiable or qualitative, the issue of the *warrant* for their inferences must be confronted. (Fielding and Fielding: 1986, 12, my emphasis)

Shortly, we will examine how qualitative or field researchers may claim, in Fielding and Fielding's terms, that they have a 'warrant for their inferences'. For the moment, however, I want to deal briefly with the argument that validity is not an issue in field research.

Validity as Unnecessary

Agar (1986) criticises 'the received view' of science, a view that centres on the systematic test of explicit hypotheses (11). This view, he argues, is inappropriate to research problems concerned with 'What is going on here?' (12) which involve learning about a world first-hand.

The implication, according to Agar, is a rejection of the standard issues of reliability and validity in favour of: 'an intensive personal involvement, an abandonment of traditional scientific control, an improvisational style to meet situations not of the researcher's making, and an ability to learn from a long series of mistakes' (12).

However, this too readily abandons any reference to the validity of the ethnographer's statements. It simply will not do to accept any account simply on the basis of the researcher's claims to 'an intensive personal involvement'. Immediacy and authenticity may be a good basis for certain kinds of journalism but ethnography must make different claims if we are to take it seriously.

Nonetheless, even a brief perusal of published articles using qualitative methods can be profoundly disturbing. When I reviewed recent volumes of two social science journals (Silverman: 1989a), I was struck by the 'anecdotal' quality of much of what I was reading. Much too frequently, the authors had fallen foul of two problems identified by Fielding and Fielding (1986):

- a tendency to select field data to fit an ideal conception (preconception) of the phenomenon
- a tendency to select field data which are conspicuous because they are exotic, at the expense of less dramatic (but possibly indicative) data (32)

As Bryman argues:

> There is a tendency towards an anecdotal approach to the use of 'data' in relation to conclusions or explanations in qualitative research. Brief conversations, snippets from unstructured interviews, or examples of a particular activity are used to provide evidence for a particular contention. There are grounds for disquiet in that the *representativeness* or generality of these fragments is rarely addressed. (Bryman: 1988, 77, my emphasis)

As already noted (in Chapter 3), another way in which field researchers have sidestepped the issue of validity is by stressing a concern to generate rather than to test theories. For instance, Glaser and Strauss' concept of 'grounded theory' (1967) seeks to generate and develop categories in order to produce delimited theories grounded in the data. While Glaser and Strauss' emphasis on 'the constant comparative method' is helpful, others have rightly criticised their apparent lack of interest in *testing* hypotheses (e.g. Fielding: 1988, 8), although, in a later work, Strauss (1987) does claim that his approach is 'designed especially for generating and testing theory' (xi).

If some field researchers sidestep the issue of validity, others reject it altogether as an appropriate issue for social research. For instance, from a feminist position, Stanley and Wise describe 'objectivity' as:

> an excuse for a power relationship every bit as obscene as the power relationship that leads women to be sexually assaulted, murdered and otherwise treated as mere objects. The assault on our minds, the removal from existence of our experiences as valid and true, is every bit as questionable. (1983, 169)

Like many feminist sociologists, Stanley and Wise argue that the validity of 'experiences' should replace supposedly male-dominated versions of 'objectivity'. Thus, although qualitative methods are held to be most appropriate for understanding women's experience, such experiences are valid or 'true' in themselves. In any event, it is argued, the goal of research is not to accumulate knowledge but to serve in the emancipation of women.

For purposes of exposition, I have chosen an extreme position – readers wanting a less dogmatic feminist approach might turn to Cain (1986). However, Stanley and Wise's argument has the merit that it reveals methodological assumptions which many feminists share.

Each assumption can be questioned as follows (for another relevant critique, see Hammersley: 1992a):

1 The assumption that 'experience' is paramount is not at all new. Indeed, it was a primary feature of nineteenth-century romantic thought (see Silverman: 1989b). As I have argued in this book (especially in Chapter 5), to focus on 'experience' alone undermines what we know about the cultural and linguistic forms which structure what we count as 'experience'.

2 Rather than being a male standard, the attempt to generate valid knowledge lies at the basis of *any* dialogue. Without the ability to choose between the truth-claims of any statement, we would be reduced to name-calling along the lines of 'you would say that, wouldn't you?'. Against certain current fashions, we ought to recognise how, when eighteenth century 'Enlightenment' thinkers pursued objectivity, they were seeking just such a way out from prejudice and unreason.

3 To assume that emancipation is the goal of research conflates yet again 'fact' and 'value'. How research is used is a value-laden, political question. The first goal of scientific research is valid knowledge. To claim otherwise is to make an alliance with an awful dynasty that includes 'Aryan science' under the Nazis, and 'Socialist science' under Stalin.

Stanley and Wise do share a common assumption with some (male) sociologists with whom they might otherwise disagree. Many qualitative researchers assume that there is a huge gulf not only between natural science and social science but between qualitative and quantitative social research. However, we must not make too much of the differences between field research and other research styles. For instance, as Hammersley (1990) points out, although replication of an ethnographic study in the same setting may be difficult, we need to understand that replication is

not always a straightforward process even in the natural sciences. Hence where research findings are not replicated this is often put down to variation in laboratory conditions and procedures (this relates to the reliability of the research instruments used – see below).

However, if social science statements are simply accounts, with no claims to validity, why should we read them? Moreover, it is paradoxical to assert that social scientists should be the only group that cannot *check* their statements: 'This is a paradoxical conclusion. While culture members freely and legitimately engage in checking claims against facts . . . the social scientist [claims to be] . . . disbarred from this on the grounds that it would "distort reality" ' (Hammersley and Atkinson: 1983, 13).

Hammersley (1990, and 1992a) has suggested that qualitative researchers *can* address issues of validity by adopting what he calls a 'subtle form of realism'. This has the following three elements:

1 Validity is identified with confidence in our knowledge but not certainty.
2 Reality is assumed to be independent of the claims that researchers make about it.
3 Reality is always viewed through particular perspectives; hence our accounts *represent* reality they do *not* reproduce it (Hammersley: 1992a, 50–51).

This is very close to Popper's account of *falsifiability* rather than verifiability as the distinguishing criterion of a scientific statement. Like Popper, Hammersley also argues that claims to validity, based on attempts at refutation, are sustained by a scientific community prepared 'to resolve disagreements by seeking common grounds of agreement' (1990, 63).

If we recognise that 'no knowledge is certain', how can we go about judging 'knowledge claims . . . in terms of their likely truth' (Hammersley: 1990, 61)? Hammersley suggests three steps:

1 The *plausibility* of the claim, given our existing knowledge.
2 The *credibility* of the claim, given the nature of the phenomena, circumstances of the research and characteristics of the researcher.
3 Where we have doubt about either 1 or 2, then we need to be convinced by the plausibility and credibility of the *evidence* (1990, 61–62).

In practice, Hammersley's points 1 and 2 create many problems. First, they exemplify the conservatism of the scientific community and the practice of 'normal science' described by Kuhn (1970). If we only accept as valid those accounts which are plausible and credible, then we are unable to be surprised and condemned to reproduce existing models of the world. Second, as we saw in the discussion of 'scientism' in Chapter 1 (p. 5) researchers' claims may sometimes be credible merely because they rely on common-sense knowledge which stands in need of explication rather than passive acceptance.

I will, therefore, stick with Hammersley's point 3. How can we be convinced by the plausibility and credibility of the evidence produced by field research? Let us review the standards by which such researchers claims to be judged.

Claims to Validity in Field Research

As I have argued, the issue of validity is appropriate whatever one's theoretical orientation or use of quantitative or qualitative data. Few contemporary social scientists have any stomach for any remaining field researchers who might maintain that our only methodological imperative is to 'hang out' and to return with 'authentic' accounts of the field.

However, I shall not discuss here many standard criteria of assessing validity, either because they are available in other methodology texts or because they are commonsensical and/or inappropriate to the theoretical logic of field research as normatively defined in Chapter 2. These criteria include:

- the impact of the researcher on the setting (the so-called 'halo' or 'Hawthorne' effect) (see Hammersley: 1990, 80–82, Landsberger: 1958)
- the values of the researcher (see Weber: 1949, and this volume, Chapter 8)
- the truth-status of a respondent's account (see this volume, Chapter 5).

Two forms of validation have been suggested as particularly appropriate to the logic of qualitative research:

1 Comparing different kinds of data (e.g. quantitative and qualitative) and different methods (e.g. observation and interviews) to see whether they corraborate one another. As already noted, this form of comparison, called *triangulation*, derives from navigation, where different bearings give the correct position of an object.
2 Taking one's findings back to the subjects being studied. Where these people verify one's findings, it is argued, one can be more confident of their validity. This method is known as *respondent validation*.

Each of these methods is discussed below where I show why I believe these methods are usually inappropriate to qualitative research.

Triangulating Data and Methods

A major early advocate of the method of triangulation is Norman Denzin (1970). This arises in the context of Denzin's discussion of the advantages and limitations of observational work. Unlike survey research, Denzin points out: 'the participant observer is not bound in his field work by prejudgements about the nature of his problem, by rigid data-gathering devices or by hypotheses' (*ibid*, 216).

However, Denzin also notes that participant observation is not without its own difficulties. First, its focus on the present may blind the observer to important events that occurred before his entry on the scene. Second, as Dalton (1959) points out, confidants or informants in a social setting may be entirely unrepresentative of the less open participants. Third, observers may change the situation just by their presence and so the decision about what role to adopt will be fateful. Finally, the observer may 'go native', identifying so much with participants that, like a child learning to talk, (s)he cannot remember how (s)he found out or articulate the principles underlying what (s)he is doing.

Given these difficulties, Denzin offers two related solutions. The first is non-contentious. It involves using multiple sources of data-collection, as part of the methodology. Thus Denzin defines participant observation: 'as a field strategy that simultaneously combines document analysis, respondent and informant interviewing, direct participation and observation and introspection' (*ibid*, 186).

Now, as an assembly of reminders about the partiality of any one context of data-collection, such a 'field strategy' makes a great deal of sense. However, it seems that Denzin wants to go beyond a recognition of the partiality of data, for his second solution to the difficulties of participant observation is to suggest that a more general practice of 'method triangulation' can serve to overcome partial views and present something like a complete picture.

As Denzin elsewhere notes, actions and accounts are 'situated'. This implies, contrary to what Denzin argues about triangulation, that methods, often drawn from different theories, cannot give us an 'objective' truth (33). So:

> multiple theories and multiple methods are . . . worth pursuing, but not for the reasons Denzin cites . . . The accuracy of a method comes from its systematic application, but rarely does the inaccuracy of one approach to the data complement the accuracies of another. (Fielding and Fielding: 1986, 35)

To counter what Fielding and Fielding rightly call Denzin's 'eclecticism' (34), they suggest that the use of triangulation should operate according to ground rules (*ibid*). Basically, these seem to operate as follows:

- begin from a theoretical perspective (e.g. interactionism)
- choose methods and data which will give you an account of structure and meaning from within that perspective (e.g. by showing the structural contexts of the interactions studied).

Even when we use a single theoretical perspective, we cannot simply aggregate data in order to arrive at an overall 'truth'. As Hammersley and Atkinson point out: 'one should not adopt a naively "optimistic" view that the aggregation of data from different sources will unproblematically add up to produce a more complete picture' (1983, 199).

The sociologist's role is not, as Dingwall (1981) reminds us, 'to adjudicate between participants' competing versions' but to understand the situated work that they do.

Of course, this does not imply that the sociologist should avoid generating data in multiple ways. As already noted, this can serve as an assembly of reminders about the situated character of action. The 'mistake' only arises in using data to adjudicate between accounts. For this reduces the role of the researcher to what Garfinkel (1967) calls an 'ironist', using one account to undercut another, while remaining blind to the sense of each account in the context in which it arises.

To conclude: the major problem with triangulation as a test of validity is that, by counterposing different contexts, it ignores the context-bound and skilful character of social interaction and assumes that members are 'cultural dopes', who need a sociologist to dispel their illusions (see Garfinkel: 1967, Bloor: 1978).

A better solution may be to distinguish 'how' from 'why' questions and to triangulate methods and data only at the 'why' stage (see Chapter 8). Equally, Dingwall (personal correspondence) has suggested that triangulation has some value where, for instance, it reveals the existence of public and private accounts of an agency's work. Here 'interview and field data can be combined . . . to make better sense of the other'. I entirely accept Dingwall's point. His example shows triangulation being used to address the *situated work* of accounts rather than, as in Denzin's case, to do *ironies*.

Exercise 7.4

This exercise is concerned with method triangulation. You should select any TWO of the methods discussed in Chapters 3–6 (i.e. observation, texts, interviews and transcripts). Then you should choose a research topic where these two methods can be applied. For example, you might want to compare your observations of a library with interviews with library-users and staff. Alternatively, you could obtain official documents about the academic aims of your university and compare these to your observations, interviews or audio-recordings of a teaching session (subject to everyone's agreement).

Now do the following:

1 Briefly analyse each of your two sources of data. What does each source tell you about your topic?
2 Identify different themes emerging in the two data sources. How far are these differences relevant for an overall understanding of the topic?
3 Using your data, assess the argument that evidence is only relevant in the context of the situation in which it arises.
4 In the light of the above, explain whether, if you had to pursue your topic further, you would use multiple methods.

Respondent Validation

Reason and Rowan (1981) criticise researchers who are fearful of contaminating their data with the experience of the subject. On the contrary, they argue, good research goes back to the subjects with the tentative results, and refines them in the light of the subjects' reactions.

This is just what Michael Bloor (1978, 1983) attempted in his research on doctors' decision-making. Bloor (1978) discusses three procedures which attempt respondent validation:

1 The researcher seeks to predict participants' classifications in actual situations of their use (see Frake: 1964).
2 The researcher prepares hypothetical cases and predicts respondents' responses to them (see also Frake: 1964).
3 The researcher provides respondents with a research report and records their reactions to it.

In his study of doctors' decision-making in adeno-tonsillectomy cases, Bloor used the third method, hoping for 'a sort of self-recognition effect' (1978, 549). Although Bloor reports that he was able to make some useful modifications as a result of the surgeons' comments, he reports many reservations. These centre around whether respondents are able to follow a report written for a sociological audience and, even if it is presented intelligibly, whether they will (or should) have any interest in it (*ibid*, 550). A further problem, noted by Abrams (1984), is that: 'overt respondent validation is only possible if the results on the analysis are compatible with the self-image of the respondents' (8).

However, Bloor concludes, this need not mean that attempts at respondents' validation have *no* value. They do generate further data which, while not validating the research report, often suggest interesting paths for further analysis (Bloor:1983, 172).

Bloor's point has been very effectively taken up by Fielding and Fielding (1986) (respondent validation is also criticised by Bryman: 1988, 78–79). The Fieldings concede that subjects being studied may have additional knowledge, especially about the context of their actions. However:

> there is no reason to assume that members have privileged status as commentators on their actions . . . such feedback cannot be taken as direct validation or refutation of the observer's inferences. Rather such processes of so-called 'validation' should be treated as yet another source of data and insight. (43)

Of course, this leaves on one side the ethics, politics and practicalities of the researcher's relation with subjects in the field (see Chapter 8). Nonetheless, these latter issues should not be *confused* with the validation of research findings.

If we reject triangulation and members' validation, how, then, are we to overcome the anecdotal quality of much field research? To answer this question, I will review what I believe to be more appropriate methods for validating studies based largely or entirely upon qualitative data.

Choosing Cases

Field research studies are usually based on one or more cases. It is unlikely that these cases will have been selected on a random basis. More likely, a case will be chosen because it allows access.

This give rise to a problem, familiar to users of quantitative methods: 'How do we know . . . how representative case study findings are of all members of the population from which the case was selected?' (Bryman: 1988, 88).

The problem of 'representativeness' is a perennial worry of case-study researchers. Let me outline a number of ways that we can address it:

Inferring from one case to a larger population: Hammersley (1992a) suggests three methods through which we can attempt to generalise from the analysis of a single case:

- obtaining information about relevant aspects of the population of cases and comparing our case to them
- using survey research on a random sample of cases
- co-ordinating several ethnographic studies.

Through such comparisons with a larger sample, we may be able to establish some sense of the representativeness of our single case.

Generalisations in terms of theories: It is important to recognise that generalising from cases to populations does not follow a purely statistical logic in field research. Quoting Mitchell (1983), Bryman thus argues that: 'the issue should be couched in terms of the generalisability of cases to *theoretical* propositions rather than to *populations* or universes' (1988, 90, my emphasis).

As an example, Bryman uses Glaser and Strauss' discussion of 'awareness contexts' in relation to dying in hospital:

> The issue of whether the particular hospital studied is 'typical' is not the critical issue; what is important is whether the experiences of dying patients are typical of the broad class of phenomena . . . to which the theory refers. Subsequent research would then focus on the validity of the proposition in other milieux (e.g. doctors' surgeries). (1988, 91)

As our understanding of social processes improves, we are increasingly able to choose cases on theoretical grounds – for instance, because the case offers a crucial test of a theory. This leads directly to the issue of how we can test hypotheses in field research.

Testing Hypotheses

Analytic Induction (AI)

The standard method of testing a hypothesis in field research is AI. Fielding (1988) notes that we should begin by defining a phenomenon and

generating some hypothesis. Then we take a small body of data (a 'case') and examine it as follows:

'[O]ne case is . . . studied to see whether the hypothesis relates to it'. If not, the hypothesis is reformulated (or the phenomenon redefined to exclude the case). While a small number of cases support 'practical certainty, negative cases disprove the explanation, which is then reformulated. Examination of cases, redefinition of the phenomenon and reformulation of hypotheses is repeated until a universal relationship is shown' (7–8).

So AI is the equivalent to the statistical testing of quantitative associations to see if they are greater than might be expected at random (random error). However: 'in qualitative analysis . . . there is no random error variance. All exceptions are eliminated by revising hypotheses until all the data fit. The result of this procedure is that statistical tests are actually *unnecessary* once the negative cases are removed' (Fielding and Fielding: 1986, 89).

An example of AI being used in a field research study will be helpful. In Bloor's study of surgeons, already discussed, he tried: 'to inductively reconstruct each specialist's own standard "decision rules" which he normally used to decide on a disposal' (Bloor: 1978, 545). These rules were then compared to each doctor's procedures for searching through relevant information.

Bloor draws upon the distinction between 'necessary' and 'sufficient' conditions for an outcome. 'Necessary' conditions are conditions without which a particular outcome is impossible. 'Sufficient' conditions are conditions which totally explain the outcome in question. For instance, a necessary condition for me to give a lecture is that I should be present at a particular time and place. Sufficient conditions may include me knowing about the subject, having my notes with me, finding an audience awaiting me, and so on. This is how Bloor reports his inductive method:

> 1 For each specialist separately, cases were provisionally classified according to the disposal-category into which they fell.
> 2 The data on all a specialist's cases in a particular disposal-category were scrutinised in order to attempt a provisional list of those case-features common to the cases in that category.
> 3 The 'deviant cases' (i.e. those cases where features common to many of the cases in the disposal-category were lacking) were scrutinised in order to ascertain whether (a) the provisional list of case-features common to a particular category could be modified as to allow the inclusion of the deviant cases; or, (b) the classificatory system could be so modified as to allow the inclusion of the deviant cases within a modified category.
> 4 Having thus produced a list of case-features common to all cases in a particular category, cases in alternative categories were scrutinised to discover which case-features were shared with cases outside the first category considered. Such shared case-features were thus judged *necessary* rather than *sufficient* for the achievement of a particular disposal.

5 From the necessary and sufficient case-features associated with a particular category of cases sharing a common disposal, the specialist's relevant decision rules were derived. (Bloor: 1978, 546, my emphasis)

This is a shortened version of Bloor's list. He adds two further stages where cases are rescrutinised for each decision rule and then the whole process is re-enacted in order to account for the disposals obtained by all the specialists in the study.

Bloor recognises that his procedure was not *wholly* inductive. Before beginning the analysis, he already had general impressions, gained from contact in the field (*ibid*, 547). We might also add that no hypothesis-testing can or should be theory-free. Necessarily, then, analytic induction depends upon both a model of how social life works (e.g. interactionism, CA, etc.) and a set of concepts specific to that model (e.g. 'frames', 'recipient-design' and, as we discuss below, the environment around 'advice-reception').

For further discussion of AI, using Bloor's study as an exemplar, see Abrams (1984).

AI may appear to be rather complicated. However, it boils down to two simple techniques:

- the search for deviant cases
- the use of the constant comparative method.

Both techniques are susceptible to simple methods of counting.

Counting in Qualitative Research

> By our pragmatic view, qualitative research does imply a commitment to field activities. It does not imply a commitment to innumeracy. (Kirk and Miller: 1986, 10)

In this part of the chapter I want to make some practical suggestions about how quantitative data can be incorporated into qualitative research. These suggestions flow from my own recent research experience in a number of studies, two of which are briefly discussed shortly.

Since the 1960s, a story has got about that no good sociologists should dirty their hands with numbers. Sometimes this story has been supported by sound critiques of the rationale underlying some quantitative analyses (Blumer: 1956, Cicourel: 1964). Even here, however, the story has been better on critique than on the development of positive, alternative strategies.

The various forms of ethnography, through which attempts are made to describe social processes, share a single defect. The critical reader is forced to ponder whether the researcher has selected only those fragments of data which support his argument. Where deviant cases are cited and explained (cf. Strong: 1979a, Heath: 1981), the reader feels more confident about the

analysis. But doubts should still remain about the persuasiveness of claims made on the basis of a few selected examples.

I do not attempt here to defend quantitative or positivistic research *per se*. I am not concerned with research designs which centre on quantitative methods and/or are indifferent to the interpretivist problem of meaning. Instead, I want to try to demonstrate some uses of quantification in research which is qualitative and interpretive in design.

I shall try to show that simple counting techniques can offer a means to survey the whole corpus of data ordinarily lost in intensive, qualitative research. Instead of taking the researcher's word for it, the reader has a chance to gain a sense of the flavour of the data as a whole. In turn, researchers are able to test and to revise their generalisations, removing nagging doubts about the accuracy of their impressions about the data.

As Cicourel (1964) noted thirty years ago, in a bureaucratic-technological society, numbers talk. Today, with sociology on trial, we cannot afford to live like hermits, blinded by global, theoretical critiques to the possible analytical and practical uses of quantification. In the mid-1990s I believe this case holds just as strongly.

In a study of oncology clinics (Silverman: 1984), I used some simple quantitative measures in order to respond to some of these problems. The aim was to demonstrate that the qualitative analysis was reasonably representative of the data as a whole. Occasionally, however, the figures revealed that the reality was not in line with my overall impressions. Consequently, the analysis was tightened and the characterisations of clinic behaviour were specified more carefully.

A major aim was to compare what, following Strong (1979a), I called the 'ceremonial order' observed in the two British National Health Service (NHS) clinics with a clinic in the private sector. My method of analysis was largely qualitative and, like him, I used extracts of what patients and doctors had said as well as offering a brief ethnography of the setting and of certain behavioural data. In addition, however, I constructed a coding form which enabled me to collate a number of crude measures of doctor and patient interactions.

My impression was that the private clinic encouraged a more 'personalised' service and allowed patients to orchestrate their care, control the agenda and obtain some 'territorial' control of the setting. In my discussion of the data, like Strong, I cite extracts from consultations to support these points, while referring to deviant cases and to the continuum of forms found in the NHS clinics.

The crude quantitative data I had recorded did not allow any real test of the major thrust of this argument. Nonetheless, it did offer a summary measure of the characteristics of the total sample which allowed closer specification of features of private and NHS clinics. In order to illustrate this, I shall briefly look at the data on consultation length, patient participation and widening of the scope of the consultation.

My overall impression was that private consultations lasted considerably longer than those held in the NHS clinics. When examined, the data indeed did show that the former were almost twice as long as the latter (20 minutes as against 11 minutes) and that the difference was statistically highly significant. However, I recalled that, for special reasons, one of the NHS clinics had abnormally short consultations. I felt a fairer comparison of consultations in the two sectors should exclude this clinic and should only compare consultations taken by a single doctor in both sectors. This sub-sample of cases revealed that the difference in length between NHS and private consultations was now reduced to an average of under 3 minutes. This was still statistically significant, although the significance was reduced. Finally, however, if I compared only *new* patients seen by the same doctor, NHS patients got 4 minutes more on average – 34 minutes as against 30 minutes in the private clinic. This last finding was not suspected and had interesting implications for the overall assessment of the individual's costs and benefits from 'going private'. It is possible, for instance, that the tighter scheduling of appointments at the private clinic may limit the amount of time that can be given to new patients.

As a further aid to comparative analysis, I measured patient participation in the form of questions and unelicited statements. Once again, a highly significant difference was found: on this measure, private patients participated much more in the consultation. However, once more taking only patients seen by the same doctor, the difference between the clinics became very small and was *not* significant. Finally, no significant difference was found in the degree to which non-medical matters (e.g. patient's work or home circumstances) were discussed in the clinics.

These quantitative data were a useful check on over-enthusiastic claims about the degree of difference between the NHS and private clinics. However, it must be remembered that my major concern was with the 'ceremonial order' of the three clinics. I had amassed a considerable number of exchanges in which doctors and patients appeared to behave in the private clinic in a manner deviant from what we know about NHS hospital consultations. The question was: would the quantitative data offer any support to my observations?

The answer was, to some extent, positive. Two quantitative measures were helpful in relation to the ceremonial order. One dealt with the extent to which the doctor fixed treatment or attendance at the patient's convenience. The second measured whether patients or doctor engaged in polite small-talk with one another about their personal or professional lives. (I called this 'social elicitation'.) As Table 7.4 shows, both these measures revealed significant differences, in the expected direction, according to the mode of payment.

Now, of course, such data could not offer proof of my claims about the different interactional forms. However, coupled with the qualitative data, they provided strong evidence of the direction of difference, as well as giving me a simple measure of the sample as a whole which contexted the few extracts of talk I was able to use. I do not deny that counting can be as

Table 7.4: *Private and NHS Clinics: Ceremonial Orders*

	Private clinic (n = 42)	NHS clinics (n = 104)
Treatment or attendance fixed at patients' convenience	15 (36%)	10 (10%)
Social elicitation	25 (60%)	31 (30%)

arbitrary as qualitative interpretation of a few fragments of data. However, providing the researcher resists the temptation to try to count everything, and bases his analysis on a sound conceptual basis linked to actors' own methods of ordering the world, then both types of data can inform the analysis of the other.

Exercise 7.5

This exercise is meant to accustom you to the advantages and limitations of simple tabulations. You should return to one of the settings which you have observed in a previous exercise.

Now follow these steps:

1 Count whatever seems to be countable in this setting (e.g. the number of people entering and leaving or engaging in certain activities).
2 Assess what this quantitative data tells you about social life in this setting. How far can what you have counted be related to any *one* social science theory or concept with which you are familiar?
3 Beginning from the theory or concept selected in step 2, indicate how you might count in terms of it rather than in terms of common-sense categories.
4 Attempt to count again on this basis. What associations can you establish?
5 Identify deviant cases (i.e. items that do not support the associations that you have established). How might you further analyse these deviant cases, using either quantitative or qualitative techniques? What light might that throw on the associations which you have identified?

Summary

I have suggested that both reliability and validity are important issues in field research. I went on to suggest that reliability can be addressed by using standardised methods to write fieldnotes and prepare transcripts. In the case of interview and textual studies, I also argued that reliability can be improved by comparing the analyses of the same data by several researchers.

I further suggested that data triangulation and member validation are usually inappropriate to validate field research. Instead, I suggested three ways of validating such research:

1 Methods of generalising to a larger population.
2 Methods of testing hypotheses.
3 The use of simple counting procedures.

I now want to conclude this chapter by an account of one further case study which illustrates several of these issues.

An Example: Analysing Advice Sequences

Heritage and Sefi (1992) (henceforth H&S) have analysed 70 instances of advice-giving sequences drawn from 8 first visits to first-time mothers by 5 different health visitors. H&S found that most advice was initiated by the professional, often prior to any clear indication that it was desired by the client.

Health visitor (HV) initiated advice took four forms:

1 Stepwise entry in the sequence below:

   (a) HV enquiry
   (b) problem-indicative response by client
   (c) request for specification by HV ('a focussing enquiry')
   (d) a specification by the client
   (e) advice-giving.

2 The same sequence but with no request for specification because the client volunteers how she dealt with the problem.
3 No client statement of how she dealt with the problem and no HV request for specification – thus stages (c) and (d) are omitted.
4 HV-initiated advice without the client giving a problem indicative response, i.e. stage (a) is followed directly by stage (e).

The majority of advice initiations analysed by H&S were of form 4. Indeed, in many cases, even the HV's enquiry was not problem-oriented but was more concerned to topicalise the issue for which advice was subsequently delivered.

The reception of advice by mothers took three forms:

1 A marked acknowledgment (MA) (e.g. 'oh right' or repeats of key components of the advice); H&S say such utterances acknowledge the informativeness and appropriateness of the advice.
2 An unmarked acknowledgment (UA) (e.g. 'mm', 'yeah', 'right' without an 'oh'). These are minimal response tokens which, H&S argue, have a primarily continuative function; they do *not* (a) acknowledge the advice-giving as newsworthy to the recipient or (b) constitute an undertaking to follow the advice and (c) can be heard as a form of resistance in themselves because, implicitly, such responses are refusing to treat the talk as advice.
3 Assertions of knowledge or competence by the mother. These indicate that the advice is redundant – hence they also may be taken as resistance.

This underlines H&S's argument about the advantages of stepwise entry into advice-giving (form 1 above). In this form of advice-giving, they find less resistance and more uptake displayed by mothers' use of marked acknowledgments. Here the HV's request for her client to specify a problem means that the advice can be recipient-designed, non-adversarial and not attribute blame.

Like H&S, in a study of AIDS counselling (Silverman *et al*: 1992), we focussed on the link between the form in which advice is delivered and its reception. Nearly all advice sequences were C-initiated and many were truncated. As in H&S's study, step-by-step sequences were more likely to produce MAs, truncated sequences usually produced UAs. The data on uptake are shown in Table 7.5.

Table 7.5: *Form of Advice and Degree of Uptake*

		Type of acknowledgement[1]	
Advice format	Number	Unmarked	Marked
P-initiated	2	0	2
C-initiated			
Step by step: full-sequence	11	1	10
Shortened	5	3	2
Truncated: no P problem elicited	32	29	3

Based on 50 advice sequences

[1] 'Unmarked' means *only* unmarked acknowledgments were given in the advice sequence; 'marked' means that at least *one* marked acknowledgment was given.

Table 7.5 shows a very clear correlation between the way in which an advice sequence is set up and the response that it elicits from the patient. In the total of 32 cases where the counsellor delivers advice without attempting to generate a perceived problem from the patient, there are only 3 cases where the patient shows any sign of uptake. Conversely, in the other 18 cases, where the advice emerges either at the request of the patient or in a step-by-step sequence, there are only 4 cases where the patient does *not* show uptake.

Table 7.5 thus shows how simple tabulations can offer a valuable means of validating impressions obtained from qualitative data-analysis.

Following the discussion above of deviant-case analysis (in Bloor's work), I now will show how the analysis of these gross findings was developed by the examination of two deviant cases. In one case a truncated sequence of generalised advice was, unusually, associated with marked acknowledgments (MAs). This is shown as Extract 7.1 below.

Extract 7.1

```
(C = counsellor; P = patient; C is talking about contracting full-blown AIDS)
 1 C:  But we can't tell you know whether uh one individual is
 2     going to or whether they're [no:t.
 3 P:                              [(It's just on proportions).=
 4 C:  That's ri:[ght.
 5 P:            [(          )
 6 C:   .hhhh A:nd obviously if someone looks after themselves they
 7     stand a better chance you know keeping fit and healthy.
 8 P:  Ye:s.
 9 C:   .hhhh The advice we give is common sense really if you think
10     about it.=To keep fit and healthy, (.)     eat a
11       [well) a balanced di:et,
12 P:    [For your natural resistance.=
13 C:   =That's [right.
14 P:           [Ye:s.
15 C:   .hh Plenty of exerci:[se:
16 P:                        [Right.
17 C:   [Uh::m   or   enough   exercise.
18 P:   [(I already get that) hheh .hhh [hhh .hhhh Too=
19 C:                                   [Yeah.
20 P:   =much of it. hhuh=
21 C:   =Enough slee:p.
22 P:   Y[es.
23 C:     [You know. All the things we should normally
24     do[: to keep healthy,
25 P:     [Right. Rather than let yourself get run down.=
26 C:   =That's ri:ght.
```

Extract 7.1 is remarkable for the large number of marked acknowledgments given by the client (lines 12, 18 and 25). How can we account for this unusual reception of truncated advice? A part of the answer seems to lie in the content of the advice given. Extract 7.1 is largely concerned with what the counsellor tells people who have a positive test-result. This leaves it open to the patient to treat what he is being told not as advice but as *information delivery* (about the advice C would give if P turned out to be seropositive).

It follows that such uptake obviously need have no direct implication for what the patient does (as opposed to what he thinks) – unlike the uptake of advice. Hence, as in Extract 7.1, P may choose to offer marked acknowledgments to what C says. But, in so doing, he may be simply showing uptake of a sequence that is hearable as information rather than personalised advice.

So when Cs formulate their talk as 'advice' but offer a generalised message (e.g. 'what we tell people who test positive'), they depart from many of the constraints of personalised advice-giving. This is because

information delivery is compatible with a wide range of response (from simple continuers to newsworthiness-tokens). Whatever the patient says will normally be heard as a receipt of information rather than, as in personalised advice sequences, bearing on the uptake of advice. Consequently, when MAs are found in such truncated advice sequences they function as strong information receipts rather than as positive uptakes of advice.

In the second deviant case, a series of questions from the counsellor led to MAs of advice even though no advice was actually tendered. Here we found that a series of hypothetical questions, increasingly specified in terms of the patient's answers, eventually led the patient to formulate the direction in which the questions were leading.

In both cases, the problematic features of response to personalised advice-giving were avoided. In the latter case, this was associated with advice-uptake but at the cost of the resources involved in the long interviews required to lead the patient in the direction desired by the counsellor. Conversely, although generalised advice sequences were not receipted as advice but as information delivery, they saved in resources by being far quicker.

I have tried to show how simple tabulations, combined with the constant comparative method and deviant case analysis, allow us to generate and to test hypotheses. However, case-study research can rarely make any claims about the representativeness of its samples. How far does this undermine its validity?

Following my earlier discussion (and the work of Mitchell: 1983), I would argue that case-study work derives its validity not from the representativeness of its samples but from the thoroughness of its analysis. For instance, while survey researchers may be satisfied with explaining 99 per cent of the variance in their samples, case-study researchers must pursue every single instance in order to refine their analysis. This I have demonstrated in my discussion of our analysis of deviant cases.

Furthermore, although we did not select a random sample, we chose our data-sets for analytical reasons in a way which tested a theoretically-derived hypothesis. So, although the HIV counselling setting differs in important respects from Heritage and Sefi's study of health visitors (see Silverman *et al*: 1992), nonetheless it provided a comparable body of advice sequences drawn from professional–client interaction. Using concepts drawn from CA, like Heritage and Sefi, we were able to support and to refine the analysis of the processes surrounding advice-reception.

Conclusion

I have concluded this chapter on rigorous field research by focussing upon the uses of simple methods of counting in largely qualitative studies. The 'advice' study uses purely descriptive statistics; the study of private practice consultations introduces some straightforward correlations. This concentration on description is not coincidental.

The kind of interpretive sociology which I have been discussing is doubly interested in description. First, like all scientific work, it is concerned with the problem of how to generate adequate descriptions of what it observes. Second, however, unlike other kinds of sociology; it is especially interested in how ordinary people observe and describe their world. Many of the procedures I have discussed here aim to offer adequate (sociological) descriptions of (lay) descriptions. Once this is recognised as the central problematic of much field research, then these procedures can be extended to what people say and write in a far broader range of contexts than the medical settings on which I have concentrated in this chapter.

More than thirty years ago, Becker and Geer (1960) recognised that adequate sociological description of social processes needs to look beyond purely qualitative methods. Everything depends, however, on the relation between the quantitative measures being used and the analytic issue being addressed: 'The usefulness of . . . statistics is a function of the theoretical problematic in which they are to be used and of the use to which they are to be put within it' (Hindess: 1973, 45).

However, I have also shown that quantitative measures are not the only way to test the validity of our propositions. Analytic induction, based upon deviant-case analysis and the constant comparative method, offers a powerful tool through which to overcome the danger of purely 'anecdotal' field research.

The time for wholesale critiques of quantitative research has passed. What we need to do now is to show the ways in which field research is every bit as rigorous as the best quantitative work.

8

The Practical Relevance of Qualitative Research

There are several claims we might like to make about the value of ethnography in policy-making. Here is one recent list, suggested by Janet Finch:

- it is relatively flexible
- it studies what people are doing in their natural context
- it is well placed to study processes as well as outcomes
- it studies meanings as well as causes (cited by Hammersley: 1992a, 125).

Together with other ethnographers, I have made similar claims both to practitioners and to research funding bodies. Unfortunately, things are not quite as easy as this list might suggest.

First, as we have already seen (especially in Chapter 2), the status of ethnography or field research as a naturalistic enterprise, concerned with meanings, is disputable. Second, as Hammersley (1992c) points out, non-ethnographic approaches can study some of these features (e.g. questionnaire panel studies can examine change over time and thus social processes). Third, as I argued in Chapter 7, the issue of the validity of qualitative research (its generalisability to larger populations, and the possible anecdotal basis of its claims) is a real one which does not exist just in the minds of policy-makers.

In responding to these problems about the practical relevance of field research, my underlying theme is simple: the relevance to practice of rigorous fieldwork informed by analytical issues rather than by social problems.

This means that it is usually necessary to refuse to allow our research topics to be defined in terms of the conceptions of 'social problems' as recognised by either professional or community groups. Ironically, by beginning from a clearly defined sociological perspective, we can later address such social problems with, I believe, considerable force and persuasiveness.

These are claims in need of demonstration. I will attempt this shortly. For the moment, however, I want to move away from the specifics of field research to review the wider debate about how all forms of sociological research stand in relation to social problems. In doing so, I shall re-state some of the arguments found in my earlier text (Silverman: 1985).