

# CHAPTER 18

---

## The analysis and interpretation of qualitative data

This chapter:

- stresses the need for a systematic analysis of qualitative data;
- emphasizes the central role of the person doing the analysis, and warns about some deficiencies of the human as analyst;
- discusses the advantages and disadvantages of using specialist computer software;
- explains the Miles and Huberman approach to analysis which concentrates on reducing the bulk of qualitative data to manageable amounts and on displaying them to help draw conclusions;
- suggests thematic coding analysis as a generally useful technique when dealing with qualitative data;
- reviews the widely used grounded theory approach;
- summarizes a range of alternative approaches; and
- finally, considers issues involved in integrating qualitative and quantitative data in multi-strategy designs.

### Introduction

Qualitative data have been described as an 'attractive nuisance' (Miles, 1979). Their attractiveness is undeniable. Words, which are by far the most common form of qualitative data, are a speciality of humans and their organizations. Narratives, accounts and other collections of words are variously described as 'rich', 'full' and 'real', and contrasted with the thin abstractions of number. Their collection is often straightforward. They lend verisimilitude to reports.

The 'nuisance' refers to the legal doctrine that if you leave an attractive object, such as an unlocked car, where children can play with it, you may be liable for any injuries they sustain.

Naive researchers may be injured by unforeseen problems with qualitative data. This can occur at the collection stage, where overload is a constant danger. But the main difficulty is in their analysis. There is no clear and universally accepted set of conventions for analysis corresponding to those observed with quantitative data. Indeed, many 'qualitative' workers would resist their development, viewing this enterprise as more of an art than a science. But for those who do wish to work within the kind of scientific framework advocated in this book, and who wish to persuade scientific or policy-making audiences, there are ways in which qualitative data can be dealt with systematically. This chapter seeks to provide an introduction to that task.

In the typology of research strategies that has been adopted in this text, the various types of flexible and multiple-strategy designs are the prime generators of large amounts of complex qualitative data.

Qualitative data are often useful in supplementing and illustrating the quantitative data obtained from an experiment or survey. Small amounts of qualitative data used as an adjunct within a largely quantitative fixed design study will not justify detailed and complex analysis. Often the need is simply to help the account 'live' and communicate to the reader through the telling quotation or apt example. However, when methods generating qualitative data form the only, or a substantial, aspect of the study, then serious and detailed attention needs to be given to the principles of their analysis.

## Two assumptions

1. *If you have a substantial amount of qualitative data you will use some kind of software package to deal with it.* Standard software, even a simple word-processing package, can do much to reduce the sheer tedium of qualitative data analysis (see Hahn, 2008 on the use of standard Microsoft Office software for a small qualitative project). For anything other than a small amount of data, the amount of drudgery you can avoid, and the ease with which you can relate to the data, make the use of a computer near to essential. There are also specialist qualitative data analysis packages which aid the process even more. See Appendix B for further details.
2. *Unless you already have experience yourself, you will be helped or advised by someone who does have experience in this type of analysis.* The dominant model for carrying out qualitative analysis has in the past been that of apprenticeship. Without accepting all the implications of such a model (which tends, for example, to include a principled inarticulacy about process) there is undoubted value in expert advice. The help provided by software is very different from that in quantitative analysis. There the 'expert's' role is largely to point you towards an appropriate test and to ensure that you understand the outcome. In qualitative data analysis, both the experienced person and the computer help you through a not very well specified process.

## Types of qualitative analysis

Box 18.1 provides a typology of possible approaches. Quasi-statistical approaches rely largely on the conversion of qualitative data into a quantitative format and have been covered under the heading of content analysis in Chapter 15. See also Abeyasekera (2005) who provides a range of suggestions.

## BOX 18.1

**Different approaches to qualitative analysis****1. Quasi-statistical approaches**

- Uses word or phrase frequencies and inter-correlations as key methods of determining the relative importance of terms and concepts.
- Typified by content analysis.

**2. Thematic coding approach**

- A generic approach not necessarily linked to a particular (or any) theoretical perspective.
- All, or parts of, the data are coded (i.e. identified as representing something of potential interest) and labelled.
- Codes with the same label are grouped together as a theme.
- Codes and themes occurring in the data can be determined inductively from reviewing the data, and/or from relevance to your research questions, previous research, or theoretical considerations.
- The themes then serve as a basis for further data analysis and interpretation.
- Makes substantial use of summaries of the themes, supplemented by matrices, network maps, flow charts and diagrams.
- Can be used on a purely descriptive or exploratory basis, or within a variety of theoretical frameworks.

**3. Grounded theory approach**

- A version of thematic coding where, as a matter of principle, the codes arise from interaction with the data.
- Codes are based on the researcher's interpretation of the meanings or patterns in the texts.
- Used to develop a theory 'grounded' in the data.
- Can be used very prescriptively following rules laid down by founders of the approach, or as a general style of analysis using a specialized terminology for different types of coding.

*Note:* There are other specialized approaches, including discourse and conversation analysis, and the analysis of narratives (i.e. stories in written, spoken or other forms). See Chapter 15, p. 361.

Thematic coding analysis is discussed in this chapter as a straightforward general approach which can be used in a wide variety of settings. The widespread popularity of grounded theory as a basis for qualitative data analysis demands its coverage in any treatment of the topic. There are a large number of other approaches, many of which call for an extensive understanding of their theoretical foundations. A brief introduction is provided in a later section of the chapter.

Whatever approach is taken, the researcher has the responsibility of describing it in detail. You have to be able to demonstrate the quality of your analysis, including how you got from the data to your interpretation.

## The importance of the quality of the analyst

The central requirement in qualitative analysis is clear thinking on the part of the analyst. Fetterman (1998) considers that the analysis is as much a test of the researcher as it is a test of the data: 'First and foremost, analysis is a test of the ability to think – to process information in a meaningful and useful manner' (p. 93). As emphasized *at the beginning of Part V*, qualitative analysis remains much closer to codified common sense than the complexities of statistical analysis of quantitative data. However, humans as 'natural analysts' have deficiencies and biases corresponding to the problems that they have as observers (see Chapter 14, p. 320). Some of these are listed in Box 18.2. Systematic, documented approaches to analysis help minimize the effects of these human deficiencies. However, there is an inescapable emphasis on *interpretation* in dealing with much qualitative data which precludes reducing the task to a defined formula. Hence, the suggestions made in this chapter are more in the nature of guidelines rather than tight prescriptions.

### BOX 18.2

#### Deficiencies of the human as analyst

1. *Data overload*. Limitations on the amount of data that can be dealt with (too much to receive, process and remember).
2. *First impressions*. Early input makes a large impression so that subsequent revision is resisted.
3. *Information availability*. Information which is difficult to get hold of gets less attention than that which is easier to obtain.
4. *Positive instances*. There is a tendency to ignore information conflicting with hypotheses already held and to emphasize information that confirms them.
5. *Internal consistency*. There is a tendency to discount the novel and unusual.
6. *Uneven reliability*. The fact that some sources are more reliable than others tends to be ignored.
7. *Missing information*. Something for which information is incomplete tends to be devalued.
8. *Revision of hypotheses*. There is a tendency either to over- or to under-react to new information.
9. *Fictional base*. The tendency to compare with a base or average when no base data is available.
10. *Confidence in judgement*. Excessive confidence is rested in one's judgement when once it is made.
11. *Co-occurrence*. Co-occurrence tends to be interpreted as strong evidence for correlation.
12. *Inconsistency*. Repeated evaluations of the same data tend to differ.

(adapted and abridged from Sadler, 1981, pp. 27–30)

## Common features of qualitative data analysis

While the possible approaches to analysis are very diverse, there are recurring features. Miles, Huberman and Saldana (2014, p. 10) give a sequential list of what they describe as ‘a fairly classic set of analytic moves’:

- giving labels (‘codes’) to chunks (words, phrases, paragraphs, or whatever), labelling them as examples of a particular ‘thing’ which may be of interest in the initial set of materials obtained from observation, interviews, documentary analysis, etc.;
- adding comments, reflections, etc. (commonly referred to as ‘memos’);
- going through the materials trying to identify similar phrases, patterns, themes, relationships, sequences, differences between sub-groups, etc.;
- using these patterns, themes, etc. to help focus further data collection; gradually elaborating a small set of generalizations that cover the consistencies you discern in the data; and
- linking these generalizations to a formalized body of knowledge in the form of constructs or theories.

This general approach forms the basis of thematic coding analysis discussed below.

## Similarity and contiguity relations

Maxwell and Miller (2008) are concerned that an emphasis on coding and categorizing is in danger of losing connections within accounts and other qualitative material. They make a distinction between *similarity relations* and *contiguity relations*. When using coding, similarities and differences are commonly used as the basis for categorization. Relationships based on contiguity involve seeing connections between things, rather than similarities or differences. We look for such relationships within a specific interview transcript or observational field notes, seeking connections between things which are close together in time or space. They can also be sought between categories and codes once they have been established as a next step in analysis.

Similar distinctions have been proposed previously, including Coffey and Atkinson’s (1996) ‘concepts and coding’ as against ‘narratives and stories’ and Weiss’s (1994) ‘issue-focused’ and ‘case-focused’ analysis.

Maxwell and Miller review several advantages of combining categorizing and connecting strategies for analysing qualitative data. They suggest that it may be useful to think in terms of:

categorizing and connecting ‘moves’ in an analysis, rather than in terms of alternative or sequential overall strategies. At each point in the analysis, one can take either a categorising step looking for similarities and differences, or a connecting step, looking for actual (contiguity based) connections between things. In fact, it is often productive to alternate between categorizing and connecting moves (p. 470).

They provide an exemplar (pp. 471–2) illustrating one way in which the two strategies can be integrated.

In their view, the ‘grounded theory’ method discussed later in the chapter actually uses this strategy although with a different terminology. In particular, Corbin and Strauss’s (2008) ‘axial

coding' is effectively the same as 'connecting analysis'. The other main approach *covered below*, 'thematic coding' analysis, while essentially based on categorizing, does not preclude following Maxwell and Miller's suggestions.

## Using the computer for qualitative data analysis

The single constant factor reported by qualitative researchers is that such studies generate very large amounts of raw data. A small ethnographic style study will generate many pages of field notes including observations, records of informal interviews, conversations and discussions. This is likely to be supplemented by copies of various documents you have had access to, notes on your own thoughts and feelings, etc., etc. A multi-method case study will produce a similar range and amount of material. Even a strictly limited grounded theory study relying solely on interviews leaves you with 20 or more tapes to be transcribed and subsequently analysed.

Before getting on with any type of analysis, you need to ensure that you know what data you have available and that they are labelled, stored and, if necessary, edited and generally cleaned up so that they are both retrievable and understandable when you carry out the analysis. A typical first analytic task of labelling or coding the materials (e.g. deciding that a particular part or segment of an interview transcript falls into the category of 'requesting information' or 'expressing doubt' or whatever) involves not only assigning that code but also having a way of seeing it alongside other data you have coded in the same way.

In the pre-computer era, these tasks were accomplished by means of file folders containing the various sources of data, markers and highlighters, and copious photocopying. One strategy was to make as many photocopies of a page as there were different codes on that page, then to file all examples of a code together. It is clear that much of the drudgery of this task can be eliminated by using a word processor. Many data sources will either be directly in the form of computer files or can be converted into them without difficulty. It may be feasible to enter field notes directly into a laptop computer. An interview tape can be entered into the word processor as it is being transcribed. Incidentally, if you have to do this yourself there is much to be said for the use of speech recognition software for this task (listen to each sentence on the tape through headphones, then repeat it out loud to activate speech recognition). It will need to be checked but modern systems can reach high standards of accuracy. Similarly, if you have access to a scanner with optical character recognition software (OCR), it is now straightforward to convert many documents into word processor files. There are some types of data for which this may not be feasible (e.g. handwritten reports).

Word processors are a boon in storing, organizing and keeping track of your data. Obviously you need to observe good housekeeping practices and should take advice on how you can survive possible hard disk crashes, loss, theft, fire, etc. Essentially, this means having multiple copies of everything, regularly kept up to date in more than one location, and in both paper and computer file versions. Word processors can also help with the coding task through 'copy' and 'paste' functions. In this way it is easy to build up files containing all instances of a particular coding whilst retaining the original file with the original data to which codes have been added.

Word processors can also be used to assist in the 'connecting' (as against categorizing) analysis of qualitative data advocated by Maxwell and Miller (2008), discussed earlier in the chapter. Marking, extracting and putting together selected data from a longer text can greatly simplify the task of data reduction needed for producing case studies, narratives, etc.

Should you go beyond using standard word processors to one of the many specialist software packages designed to help with qualitative data analysis?

## Using specialist qualitative data analysis (QDA) packages

There are many computer packages specifically designed for researchers to use when analysing qualitative data (commonly referred to as CAQDAS – computer-assisted qualitative data analysis). The most widely used has probably been NUD\*IST (Non-numerical, Unstructured Data Indexing, Searching and Theorizing), a catchy acronym which encapsulates the central features of many of the packages – indexing, searching and theorizing. NUD\*IST has now been superseded by NVivo, developed by the same organization, QSR International (<http://www.qsrinternational.com>). It can be used profitably in most situations where you have substantial amounts of qualitative data, and for many different types of study, including grounded theory, conversation and discourse analysis, ethnographic studies, phenomenological studies, action research, case studies, and mixed method research. If you have facility in its use it is also a valuable tool when carrying out literature reviews.

While NVivo is the preferred option for qualitative data analysis in many institutions and hence is likely to be readily available and to receive support, there are several other packages worth considering for particular situations or types of data – see Appendix B for details.

When deciding whether or not to use specialist software, the advantages of time-saving and efficiency when analysing large amounts of data (once you have gained familiarity with a package), should be weighed against the time and effort taken to gain that familiarity. Box 18.3 lists some general advantages and disadvantages in their use. García-Horta and Guerra-Ramos (2009) discuss the use of two different packages with interview data, concluding that 'CAQDAS is of great help and can enhance interview data analysis; however, careful and critical assessment of computer packages is encouraged. Their capabilities must not be overestimated, since computers are still unable to perform an independent rational process or substitute the analyst's capacities' (p. 151).

Richards (2002), the prime mover in the development of the NUD\*IST and NVivo packages, expresses concerns that the full potential of computer-based analysis is not being realized. More seriously, the packages may actually be having negative effects. Because the coding and sorting tasks can be carried out more effectively and efficiently using a computer package, users tend to focus excessively on this aspect:

The code-and-retrieve techniques most easily supported by computers and most demanded by users are techniques most researchers had used at some time, for sorting out the mess of complex data records. But they were not much discussed in the literature before computing, and not at all clearly associated with the goal of theorizing common to most qualitative methodologies. So computing became associated with techniques that are generic, easily learnt and that emphasize data management, and description. Significantly, these are aspects of practical research ignored or even spurned by theoretical writers (p. 266).

It could be argued that users are simply replicating their previous paper-based, cut and paste, highlighter employing, practices on the computer, but it is undoubtedly true that packages are capable of much more than this; for example, they can include tools for doing more

## BOX 18.3

## Advantages and disadvantages of specialist QDA packages

### Advantages

- They provide an organized single location storage system for all stored material (also true of word processing programs).
- They give quick and easy access to coded material (e.g. examples of a particular theme) without using 'cut and paste' techniques.
- They can handle large amounts of data very quickly.
- They force detailed consideration of all text in the database on a line-by-line (or similar) basis.
- They help the development of consistent coding schemes.
- They can analyse differences, similarities and relationships between coded elements.
- Many have a range of ways of displaying results.

### Disadvantages

- Proficiency in their use takes time and effort.
- There may be difficulties in changing, or reluctance to change, categories of information once they have been established.
- Particular programs tend to impose specific approaches to data analysis (depends on the program – see Appendix B).
- Tendency to think that simply because you have used specialist software you have carried out a worthwhile analysis. A focus on coding and other technical aspects can give less emphasis to interpretation.

interpretation once the coding is done. Encouragingly, theory-building software has in recent years been developed to such an extent that it is probably the most widely used type. While, as pointed out by Maxwell and Miller (2008), most of these uses have been based on a prior categorizing analysis, many of the current programs allow the user to create links among and between *any* segments, both within and between contexts, and to display the resulting networks.

## Dealing with the quantity of qualitative data

Qualitative data can easily become overwhelming, even in small projects. Hence you need to find ways of keeping it manageable. This process starts before any data are collected when you focus the study and make sampling decisions about people to interview, places to visit, etc. During and after data collection you have to reduce the data mountain through the production of summaries and abstracts, writing memos, etc. Miles *et al.* (2014, p. 12) refer to this as *data condensation*. They emphasize that this is a part of analysis and not a separate activity. Decisions about what to select and to summarize, and how this is then to be organized, are analytic choices.



## Good housekeeping

Even a small project producing qualitative data can easily leave you overwhelmed with lots of pieces of information of many different types. Possible ways of keeping track include the use of:

- *Session summary sheets.* Shortly after a data collection session (e.g. an interview or observation session) has taken place and the data have been processed, a single sheet should be prepared which summarizes what has been obtained. It is helpful if this sheet is in the form of answers to summarizing and focusing questions. These might include who was involved, what issues were covered, what is the relevance to your research questions (effectively what was the purpose of the session), new questions suggested and implications for subsequent data collection.
- *Document sheets.* A similar sheet prepared for each document collected. This clarifies its context and significance, as well as summarizing the content of lengthy documents. The session summary and document sheets assist in data reduction, an important part of the analysis process.
- *Memoing.* A memo can be anything that occurs to you during the project and its analysis. Memoing is a useful means of capturing ideas, views and intuitions at all stages of the data analysis process.
- *The interim summary.* This is an attempt to summarize what you have found out so far and highlight what still needs to be found out. It is recommended that this is done before you are half-way through the time you have available for data collection. The summary should cover not only what is known but also the confidence you have in that knowledge, so that gaps and deficiencies can be spotted and remedied. *Flexible designs enable you to do this in a way which would not be feasible in a fixed design study but to capitalize on this flexibility you must force yourself to find the time to do this interim summary while you can still take advantage of its findings to direct and focus the later phases of data collection.* The summary can also usefully incorporate a *data accounting sheet* which lists the different research questions and shows, for different informants, materials, settings, etc., whether adequate data concerning each of the questions have been collected.

## Thematic coding analysis

Thematic coding analysis is presented here as a generic approach to the analysis of qualitative data. It can be used as a realist method, which reports experiences, meanings and the reality of participants or as a constructionist method, which examines the ways in which events, realities, meanings, and experiences are the effects of a range of discourses operating within society.

Coding has a central role in qualitative analysis. Gibbs (2007), in a very clear and accessible discussion, introduces it as follows:

Coding is how you define what the data you are analyzing are about. It involves identifying and recording one or more passages of text or other data items such as the parts of pictures that, in some sense, exemplify the same theoretical or descriptive idea. Usually, several passages are identified and they are then linked with a name for that idea – the code. Thus all the text and so on that is about the same thing or exemplifies the same thing is coded to the same name (p. 38).