

Ensuring rigour in qualitative research

CLIVE SEALE, DAVID SILVERMAN *

The issues of reliability and validity in qualitative research are not as readily codified as has been the case for quantitative research. However, a variety of methods do exist and are reviewed here. This general review is followed by a detailed illustration of selected techniques, including the use of counting in qualitative research, the development of systematic coding schemes with the aid of computer programmes, searching for deviant cases and the use of the transcription techniques of conversation analysis. The examples given are drawn from a variety of studies conducted by the authors.

Key words: methodology, public health research, qualitative methods, reliability, validity

This paper outlines methods for improving the rigour of qualitative research, using examples from our own research. Particular attention is paid to the role of counting, the use of computer programmes for data analysis and to the use of transcription techniques developed for conversation analysis which improve the accuracy with which data is recorded.

There is now a considerable body of knowledge and experience about the enhancement of the reliability and validity of qualitative methods, doubts about which have long been the main reason for suspicion of qualitative approaches. Svensson¹ was correct to say that the recipe-like standardization that one sees in textbooks devoted to quantitative methods is unavailable in relation to qualitative methods. Whether such simplification is desirable in either field is, frankly, questionable. In particular, it sidesteps the way in which theoretical assumptions underlie the interpretation of both qualitative and quantitative data.² However, more can be done to publicize the techniques now available to the qualitative researcher and this article is an attempt to do this.

A COMPARISON OF QUANTITATIVE AND QUALITATIVE APPROACHES

Quantitative and qualitative researchers have often diverged in their approaches to reliability and validity. This has influenced the choice of technique for gathering data. For example, in quantitative health research, observation is not generally seen as a very important method of data collection. This is because it is difficult to conduct observational studies on large samples. Quantitative researchers also argue that observation is not a very reliable data collection method because different observers may record different observations. If used at all, observation is held to be appropriate only at a preliminary or exploratory stage of research.

Conversely, observational studies have been fundamental to much qualitative research. Starting with the pioneering case studies of non-Western societies by early anthropologists^{3,4} and continuing with the work by sociologists in Chicago prior to the Second World War,⁵ the observational method has often been the chosen method to understand health behaviours at first hand.⁶⁻¹¹

These contrasts are also apparent in the treatment of texts and documents. Quantitative researchers try to analyse written material in a way which will produce reliable evidence about a large sample. Their favoured method is content analysis in which the researchers establish a set of categories and then count the number of instances that fall into each category. The crucial requirement is that the categories are sufficiently precise to enable different coders to arrive at the same results when the same body of material (e.g. newspaper headlines) is examined.¹²

In qualitative research, small numbers of texts and documents may be analysed for a very different purpose. For instance Armstrong¹³ and Martin¹⁴ have shown how medical texts have redefined the patient and the body before the doctor's gaze. Another aim is to understand how participants' categories are used in concrete activities such as telling stories,^{15,16} assembling files^{17,18} or describing family life.¹⁹ The reliability of the analysis is less frequently addressed. Instead, qualitative researchers make claims about their ability to reveal the local practices through which given end-products (stories, files and descriptions) are assembled.

Interviews are commonly used in both methodologies. Quantitative researchers administer interviews or questionnaires to random samples of the population. Fixed-choice questions (e.g. 'yes' or 'no') may be preferred because the answers they produce lend themselves to simple tabulation, unlike open-ended questions which produce answers which need to be subsequently coded. A central methodological issue for quantitative researchers is the reliability of the interview schedule and the representativeness of the sample.

Authenticity rather than reliability is often the issue in qualitative research. The aim is usually to gather an

* C. Seale¹, D. Silverman¹

¹ Department of Sociology, Goldsmiths College, London, UK

Correspondence: Clive Seale, PhD, Department of Sociology, Goldsmiths College, Lewisham Way, London SE14 6NW, UK, tel. +44 171 9197729, fax +44 171 9197713

'authentic' understanding of people's experiences and it is believed that open-ended questions are the most effective route towards this end. So, for instance, in gathering life-histories or in interviewing the parents of handicapped children²⁰ people may simply be asked: 'tell me your story'. Qualitative interview studies are often conducted with small samples and the interviewer-interviewee relationship may be defined in political rather than scientific terms.²¹

Finally, transcripts of audio recordings are rarely used in quantitative research, probably because of the assumption that they are difficult to quantify. However, following Scheff,²² we would argue that "intensive verbatim records of single cases ... can generate data-driven theories" (p. 74). Hence, audio recordings are an increasingly important part of qualitative health research.^{23,24} Transcripts of such recordings, based on standardized conventions, provide an excellent record of naturally occurring interactions. Compared to field notes of observational data, recordings and transcripts can offer a highly reliable record to which researchers can return as they develop new hypotheses.

RELIABILITY AND VALIDITY IN QUALITATIVE RESEARCH

It is sometimes suggested that the assessment of the quality of qualitative data should transcend conventional standards. The quality of qualitative research, it is argued "cannot be determined by following prescribed formulas. Rather its quality lies in the power of its language to display a picture of the world in which we discover something about ourselves and our common humanity". (p. 119).²⁵

We believe that such a position can amount to methodological anarchy and resist this on 2 grounds. First, it simply makes no sense to argue that all knowledge and feelings are of equal weight and value. Even in everyday life, we readily sort fact from fancy. Why, therefore, should science be any different? Second, methodological anarchy offers a clearly negative message to the audiences of qualitative health research, suggesting that its proponents have given up claims to validity.

Hence, we argue that rigour and validity are properly important in qualitative health research. A variety of methods have been developed to achieve these in qualitative research and these are listed below, with references given to studies employing them or key discussions in the literature. In the sections that follow this, we will illustrate some of these strategies in more detail.

- Supporting generalizations by counts of events (quasi-statistics).²⁶ This can address a common concern about the reporting of qualitative data – that anecdotes supporting the writer's argument have been selected, or that undue attention has been paid to rare events, at the expense of more common ones.
- Ensuring representativeness of cases, including the use of combined qualitative and quantitative methods to support generalizations²⁷ and recognition of the merits of representative (often random) sampling as well as

theoretical sampling (the selection of cases according to theoretical criteria).²⁸

- Testing hypotheses in data analysis, including consideration of deviant case analysis (instances in data which contradict emerging hypotheses²⁹), analytic induction³⁰ and grounded theory.²⁸ These are approaches for analysing and collecting qualitative data that ensure that generalizations are supported by adequate evidence.
- Using computer programmes to assist qualitative data analysis, thus ensuring systematic analysis of representative instances of data.³¹
- Recording data objectively and comprehensibly, including the use of audiotapes, videotapes and different levels of detail in the transcription of data.²

None of these techniques can be said to solve all the problems of reliability and validity finally but used with due regard to their limitations can advance these causes in research reports. Examples from our own work illustrate the use of some of these techniques.

Counting, computers and deviant case analysis

A common concern in the reporting of qualitative data is that of anecdotalism, raising the charge of implausibility. The selection of examples that support the writer's preferences is evident, for example, in Graham and Oakley's³² observational study of the relationships between mothers and doctors in antenatal clinics. The report consists of a series of atrocity stories about insensitive behaviour by doctors, contrasted with mothers whom the researchers believe to be already 'knowledgeable about pregnancy and birth'. The reader is not reassured that the researchers have tried to seek out instances in their data showing sensitive treatment by doctors or of mothers who hope to gain knowledge from the doctors because their own is inadequate (in other words, deviant case analysis).

Presenting simple counts of events can help readers gain a sense of how representative and widespread particular instances are. This was shown in a study of 163 elderly people living alone in their last year of life, where relatives and others were interviewed after the deaths of the people concerned.³³ Here is an extract from the report of this study.

It was very common for the people living on their own to be described either as not seeking help for problems that they had (65 instances covering 48 people), or refusing help when offered (144 instances in 83 people). Accounts of this often stressed that this reflected on the character of the person involved, although other associations were also made. In particular, 33 speakers gave 44 instances where they stressed the independence which this indicated: '(She) never really talked about her problems, was very independent...'; '(She) was just one of those independent people who would struggle on. She wouldn't ask on her own'; 'She used to shout at me because I was doing things for her. She didn't like to be helped. She was very independent. Being 'self sufficient', 'would not be beaten', and being said to 'hate to give in' were associated with resisting help.(p. 84)³³

Reading through all of the interviews and marking each instance of talk about help was the first step towards generating this report. Using a computer programme (ETHNOGRAPH) to select all these segments then

made a further coding exercise possible, in which instances of talk about not seeking help were distinguished from talk about refusing help when offered. A variety of reasons for these 2 events were also coded, distinguishing the 33 people who understood independence to be indicated by this behaviour. The report then goes on to show how other meanings were conferred on these events by speakers (for example, that the elderly person was 'being difficult' rather than being independent) but that these were less common sentiments.

A systematic coding scheme makes it possible to conduct deviant case analysis. On the one hand, this can reassure about anecdotalism, but it can also help in developing more inclusive theories to account for the data. This is shown in a study of 149 people who died alone, where once again the data was gathered in interviews with surviving friends, neighbours and relatives.³⁴ The majority of the respondents said that dying alone had been regrettable and they would like to have been present. Much of the research report concerned the ways respondents sought to repair the damage to their moral reputation caused by being absent at the time of death. However, the systematic coding procedure revealed 5 instances that deviated from this pattern. Here are 2 of these.

(The son of an 83 year old woman) said that 'Death is a private thing' when asked why he had not wanted to be there. This was associated with a view that it was beneficial to be unaware of unpleasant events, turning from this idea to the subject of his mother's unawareness of her dying: 'You've nothing to worry about if you don't know.'... Additionally, this speaker mentioned early in the interview that his wife suffered from 'early senile dementia so I can't work'. Clearly this information had affected the interviewer, as she wrote at the end: 'Respondent's wife suffering from premature senility but never mentioned as causing problems re: (care of) mother'. The speaker had provided the interviewer with sufficient reason to absolve him from any blame.(p. 387)³⁴

(The wife of a publican) was quite explicit that her reason for wanting to have been there at the time of death had nothing to do with the ideal of accompaniment. Her husband's long standing alcoholism had left her feeling little warmth for him; her reason for wishing she had been there was to save her employee the distress of finding the body.(p. 388)³⁴

These respondents are clearly seeking to demonstrate their moral adequacy by alternative means. The search for deviant cases on this occasion served to deepen an understanding of the ways in which people seek to defend moral reputations and to increase the plausibility of the final research report.

Having considered the role of counting and the search for deviant cases, we now turn to another aspect of rigour in qualitative research: the use of detailed transcription techniques.

The use of detailed transcription techniques

A recent study³⁵ involving one of the authors (DS) sought to examine the basis upon which interpretive judgements were made about the content of a series of audiotaped doctor-patient interviews between 3 oncologists and their newly referred cancer patients. It was during this

interview that the patients were supposedly informed that their cancer was incurable.

Two independent transcriptions were performed. In the first, an attempt was made to transcribe the talk 'verbatim' i.e. without grammatical or other tidying up. The second transcription was informed by the analytic ideas and transcription symbols of conversation analysis.³⁶ This provided additional information on how the parties organized their talk and, we believe, represents a more objective, comprehensive and, therefore, more reliable recording of the data because of the level of detail given by this method.

Using the first transcription, 3 independent coders, who had been trained to be consistent, coded the same material. The intercoder reliability was then estimated. The inconsistencies between the coders may have reflected some ambiguity in the data, some overlap between the coding categories or simple coding errors. Concerning a single topic on which such reliability was judged, in 21 of the 22 interviews all 3 coders agreed that the patient had been told that the condition was incurable. For one case, however, there was a discrepancy involving the interpretation of the interview data. The case concerned a man aged 64 years at entry to study, who had been diagnosed with small-cell lung cancer one month prior to the interview, the disease having spread by the time of the initial diagnosis. This patient died almost exactly 12 months after the diagnosis.

On reading the first transcript 2 of the 3 coders believed that the patient had not been told that the condition was incurable. The following is a transcript of the relevant text.

Example 1

(D, doctor; P, patient; W, wife)

D: Its very hard to be absolutely dogmatic about any predictions with these things. But despite all of those things, in the majority of people the disease does come back, even from the beginning.

P: Yes,

D: And if it does come back, we can try other drugs which may control it for a little while, but generally all that you can try and do is control the symptoms.

P: Yes Mm.

D: Uhm, the first time gives us the best chance for a longer survival, hopefully long term, but the odds are generally against that. But if [our emphasis] we do nothing for these sorts of diseases, it kills you within a couple of months.

W: The breakdown could be that quick could it?

P: Mmm

D: Well actually its quite spectacularly fast.

The discrepancy appears to have occurred primarily because the word 'incurable' was not used by the doctor. In addition, one of the 2 coders believed that there was some ambiguity attached to the doctor's statement in relation to the cancer.

The coder believed that the doctor's use of the word 'if' (in his third statement) was responsible for this.

Given this initial transcription, we can only clarify this ambiguity by observing that the wife's comment seems to reveal an understanding of the seriousness of the disease.

Can a retranscription, according to the more exacting standards required by conversation analysis, tell us how this doctor and patient heard each other's talk?

We might initially think that this is a fruitless path to follow. How could a transcription of talk give us entry into people's minds and show us what people are thinking? The answer lies in what conversation analysis reveals about what people are doing in their talk. Unlike the earlier extracts, transcripts which follow the conventions of conversation analysis do not attempt to 'clean up' such features of spoken English as hesitations ('er'), breaths ('hh'), non-dictionary words ('yer') or pauses. Here is the second transcription of the relevant passage, beginning just under 2 lines after the first extract begins and ending before the patient's wife's interjection

Example 2

- 1 D: But er despite all of those things, in the majority
2 of people the disease does come back
3 (0.8)
4 D: even from the beginning.
5 P: Yes
6 D: And: (0.4) if it does come back we can try
7 other drugs which may control it for a little while
8 P: mm um
9 D: but generally all that you can try and do is control
10 the symptoms.
11 P: Yes mm.
12 D: Uhm, the first time gives us the best chance
13 for a longer (0.5) survival hopefully long term
14 P: hhm
15 D: but the odds are generally against that.
16 P: Yes um (0.4)
17 D: But if we do nothing for these sorts of diseases it
18 kills you within a couple of months.
19 P: Yes

After the doctor has offered a statement which approximates to incurability ('in the majority of people the disease does come back'), the continued retranscription (line 3) reveals a 0.8 s slot which the patient does not use to indicate his receipt of this information. Now it appears that the patient can be heard actively avoiding receiving the category 'incurability'.

Presumably, the doctor monitors the patient's (lack of) response in the same way because he continues by upgrading his statement ('even from the beginning'). Now, if the patient were still to say nothing, his silence would be distinctly noticeable possibly indicating (to the doctor) some psychological category such as 'denial'.

However, at this point, the patient does provide a response token ('yes', line 5). Its presence here, immediately after the doctor has conveyed the bad news, and in an upgraded form ('yes rather than 'mm') strongly suggests that the patient has received the seriousness of his condition.

Both this 'yes' and the patient's 'yes mm' after further bad news (line 11) are available in the first transcript. However, the retranscription gives additional evidence of the patient's receipt of what the doctor is saying about his incurability. On lines 16 and 19, the patient is heard to say 'yes' after the doctor has said that 'the odds are

generally against (long-term survival)' and that 'if we do nothing for these sorts of diseases it kills you within a couple of months.' So we can now say, with some confidence, that the patient, as well as his wife, receives the bad prognosis.

To recapitulate, our aim was to determine whether we could improve the reliability and validity of qualitative analysis. First, we sought to identify and to explain the differences between the trained coders using rough transcripts. Next, drawing upon the transcription symbols and concepts of conversation analysis, we sought to reveal subtle features in the talk, showing how both the doctor and patients produced and received audible ambiguities in the patient's prognosis. This involved a shift of focus from the coders' readings to how participants demonstrably monitor each other's talk. Once we pay attention to such detail, judgements can be made that are more convincingly valid. Inevitably, this leads to a resolution of the problem of intercoder reliability.

Let us briefly look at one further example where it is possible to show how a medical professional monitors a client's communication. Example 3 below is taken from a British hospital centre where a patient (P) is receiving mandatory counselling from a health adviser (C) in order to achieve informed consent to an HIV test.³⁷

Example 3

- 1 C: so you know it's not hh *dead* set on ten years hh now
2 there are other people who could be HIV positive but not
3 actually develop AIDS as such hh so they could be (.)
4 carriers they could (.) stay well hh but pass the virus
5 to people that they have sex with hh this is why we say
6 hh if you don't know the person that you're with (0.6)
7 and you're going to have sex with them hh it's *important*
8 that you tell them to (0.3) use a condom.
9 (0.8)
10 C: or to practice safe sex that's what using a condom means.
11 (1.5)
12 C: okay?
13 (0.3)
14 P: uhum
15 (2.4)
16 C: has your pa:rtner ever used a condom with you?
17 (1.0)
18 P: n:o
19 (1.5)
20 C: do yer know what a condom looks like?
21 (0.5)
22 P: (I don't)
23 (0.3)
24 C: (Did you-) (0.3) have you perhaps- (1.0) a condom shown
25 to you (.) at school? = or:?
26 P: no:
27 C: yer didn't alright, =okay hhh
28 (2.0)
29 C: is there anything that yer worried about in terms of yer
30 test if it's done today? (.) would you like the test
31 first of all to be done?

In example 3 above, on lines 1–8, the health adviser offers an advice package about the difficulty of identifying risky partners who nonetheless appear well. This leads into advice about condom use (lines 5–8). Soon afterwards (line 16 onwards), although the topic stays the same, the

health adviser moves away from delivering advice to asking questions of the patient. By line 29, the health adviser has shifted entirely away from the topic of condom use.

The question we want to ask is: how can we relate this shift of format (from advice to questioning) and topic to how the health adviser might have heard the patient's earlier contributions?

First, we note that at line 10, the health adviser's explanation of what using a condom means follows directly after a 0.8 s pause on line 9. Unlike the 0.3 s pause on line 8, this pause occurs at an audible turn transition point. Thus, the 0.8 s pause can be monitored as a pause that specifically 'belongs' to the patient, where such pauses may be heard to indicate inattention, lack of interest or similar. The health adviser's explanation at line 10 demonstrates that she monitors the pause on line 9 in exactly this way. Line 10 thus unmistakably stands as a remedy to an obvious problem that the health adviser hears in the conversation.

In the same way, the health adviser's 'okay?' on line 12 is a further remedy to exactly the same problem – here expressed by the patient's non-response at the very next possible turn transition point (in the 1.5 s pause on line 11). Now, finally, the patient responds with 'uhm' (line 14). However, note that, despite the 2.4 s slot (line 15) available, she does not enlarge on this minimal utterance. We now see how the health adviser demonstrates how she monitors the patient's pauses through switching out of advice giving and into questioning the patient about her familiarity with condoms (lines 16–25) and, why, when it turns out that the patient knows nothing at all about them, the health adviser changes topic on line 29. This topic change, as we have seen, follows the patient's failure to respond at all to the health adviser's initial information and advice about condoms (lines 9 and 11) and her negative responses to the health adviser's questions about the patient's knowledge (lines 18, 22 and 26). Now the health adviser demonstrates that she has monitored the pointlessness of pursuing the topic of condoms with a markedly non-aligned client by switching topics at line 29. In short, we have no need to speculate about what the patient means by her pauses or what the health adviser is thinking during them. Rather, the trajectory of the health adviser's talk between lines 11 and 31 reveals how she is demonstrably monitoring the patient's communication. Consequently, we do not need to appeal to interrater agreement, as the reliability and validity of the analysis is based on evidence about what the professional and client can themselves be shown to be doing. (For a review of the literature on how pauses between turns can be heard between speakers see Heritage,³⁶ pp. 265–80.)

We offer these examples to stress the more general point that we wish to make: that techniques and approaches exist that can help to improve the reliability and validity of qualitative data and of research reports based on these data. These are not always packaged up in the neat – some might say too neat – prescriptions that one finds in manuals for conducting quantitative studies, but we hope to have given readers a guide to some of the key discussions of these topics.

REFERENCES

- 1 Svensson P-G. Qualitative methodology in public health research. *Eur J Public Hlth* 1995;5(2):71.
- 2 Silverman D. *Interpreting qualitative data: methods for analysing talk, text and interaction*. London: Sage, 1993.
- 3 Malinowski B. *Argonauts of the Western Pacific*. London: Routledge, 1922.
- 4 Radcliffe-Brown AR. *The Andaman Islanders*. Glencoe, Ill: Free Press, 1948.
- 5 Thomas W, Znaniecki F. *The Polish peasant in Europe and America*. New York: Alfred Knopf, 1927.
- 6 Goffman E. *Asylums*. New York: Doubleday Anchor, 1961.
- 7 Strong P. *The ceremonial order of the clinic*. London: Routledge, 1979.
- 8 Silverman D. *Communication and medical practice: social relations in the clinic*. London: Sage, 1987.
- 9 Atkinson P. *Medical talk and medical work: the liturgy of the clinic*. London: Sage, 1995.
- 10 Prior L. Policing the dead: a sociology of the mortuary. *Sociology* 1987;21(3):335-76.
- 11 Stimson G. Place and space in sociological fieldwork. *Social Rev* 1986;34(3):641-56.
- 12 Berelson B. *Content analysis in communicative research*. New York: Free Press, 1952.
- 13 Armstrong D. *The political anatomy of the body: medical knowledge in Britain in the twentieth century*. Cambridge: Cambridge University Press, 1983.
- 14 Martin E. *The woman in the body: a cultural analysis of reproduction*. Buckingham: Open University Press, 1989.
- 15 Propp VI. *Morphology of the folktale*. Austin and London: University of Texas Press, 1968.
- 16 Sacks H. On the analyzability of stories by children. In: Turner R, editor. *Ethnomethodology*. Harmondsworth: Penguin, 1974:216-32.
- 17 Cicourel A. *The social organization of juvenile justice*. New York: John Wiley, 1968.
- 18 Gubrium J, Buckholdt D. *Describing care: image and practice in rehabilitation*. Cambridge, MA: Oelschlagel, Gunn & Hain, 1982.
- 19 Gubrium J. *Out of control: family therapy and domestic disorder*. London: Sage, 1992.
- 20 Baruch G. Moral tales: parents' stories of encounters with the health profession. *Sociol Hlth Illness* 1981;3(3):275-96.
- 21 Finch J. 'It's great to have someone to talk to': ethics and politics of interviewing women. In: Bell C, Roberts H, editors. *Social researching: politics, problems, practice*. London: Routledge, 1984:70-87.
- 22 Scheff T. Single case analysis in the health sciences. *Eur J Public Hlth* 1995;5(2):72-4.
- 23 Drew P, Heritage JC, editors. *Talk at work*. Cambridge: Cambridge University Press, 1992.
- 24 Maynard DW. Interaction and asymmetry in clinical discourse. *Am J Sociol* 1991;97(2):448-95.
- 25 Buchanan DR. An uneasy alliance: combining qualitative and quantitative research methods. *Hlth Educat Q* 1992;Spring:117-35.
- 26 Silverman D. Going private: ceremonial forms in a private oncology clinic. *Sociology* 1984;18(2):191-202.
- 27 Bryman A. *Quantity and quality in social research*. London: Unwin Hyman, 1988.
- 28 Glaser BG, Strauss AL. *The discovery of grounded theory: strategies for qualitative research*. Chicago: Aldine, 1967.
- 29 Heath C. The opening sequence in doctor-patient interaction. In: Atkinson P, Heath C, editors. *Medical work: realities and routines*. Basingstoke: Gower, 1981:71-90.

CONCLUSION

What all these analyses suggest is that continuing and detailed attention needs to be paid to the quality of qualitative data and the ways in which judgements are made about its content.

30 Bloor M. On the analysis of observational data: a discussion of the worth and uses of inductive techniques and respondent validation. *Sociology* 1978;12(3):545-57.

31 Weitzman E, Miles M. Computer programmes for qualitative data analysis. London: Sage, 1994.

32 Graham H, Oakley A. Competing ideologies of reproduction: medical and maternal perspectives on pregnancy. In: Roberts H, editor. *Women, health and reproduction*. London: Routledge and Kegan Paul, 1981:50-72.

33 Seale CF. Living alone towards the end of life. *Ageing Soc* 1996;16:75-91.

34 Seale CF. Dying alone. *Social Health Illness* 1995;17(3):376-92.

35 Clavarino A, Najman J, Silverman D. Assessing the quality of qualitative data. *Qualitative Inquiry* 1995;1(2):223-42.

36 Heritage J. *Garfinkel and ethnomethodology*. Cambridge: Polity, 1984.

37 Silverman D. *Discourses of counselling: HIV counselling as social interaction*. London: Sage, 1996.

Received 9 January 1996, accepted 12 November 1996