Thank you for downloading this document from the RMIT Research Repository.

The RMIT Research Repository is an open access database showcasing the research outputs of RMIT University researchers.


Citation:

See this record in the RMIT Research Repository at:

Version: Accepted Manuscript

Copyright Statement: © Emerald Group Publishing

Link to Published Version:
N/A
Missing Data

Finding “Central” Themes in Qualitative Research

By

Supriya Singh and Lyn Richards

Assoc Prof. Supriya Singh
Senior Research Fellow
Faculty of Business
RMIT University
Box 2476V
Melbourne 3001, Australia

Lyn Richards
Director, Research Services,
Qualitative Solutions and Research,
Second floor, 651 Doncaster Rd.,
Doncaster, Vic 3108, Australia

Email: supriya.singh@rmit.edu.au
Ph: +61 3 9925 1327
Fax: +61 3 9925 1313

Email: lyn@qsr.com.au
Ph) +61 3 9840-1100
Fax: +61 3 9840-1500
Missing Data: Finding “Central” Themes in Qualitative Research

Abstract
Qualitative sociology relies heavily on the "emergence" of themes, construction of categories "out of" data, and linking those categories to form theories. But the processes by which the researcher constructs theory are rarely reported in detail. The literature is clear on the goals of theme emergence and theory construction, but not nearly so clear on the ways their relationships to data are established, or the processes by which categories can be linked to and tested through, let alone "grounded" in, data. Researchers can remain untroubled by these issues, and reports often do not mention doubts or dilemmas attending the process of theory construction.

This paper offers a contribution to the practical consideration of theory construction processes, and particularly to the practice of exploring the complex and uneven data records that usually confront the researcher. It argues that "core" categories are quite normally not evident in data, and that the gaps in data may be more significant than areas that are well documented. We report two experiences in unrelated projects of two kinds of “missing data”. The first relates to the emergence of grounded theory, in which the categories that proved central to the analysis were hardly represented in the "raw" data. We trail the processes by which the researchers became increasingly convinced of the centrality of these "missing" themes, and set about establishing an explanation for their failure to appear in the open-ended accounts that formed the data. We then raise a question not often asked: how to establish the validity of the explanation for which there is at first sight no evidence? In the second type of missing data, the data are missing, because the project raised questions at a late stage which the data did not directly address. Changes in the research question are common. However when the question emerges towards the end, especially of a funded project, then it becomes the starting point of a new project.

Introduction
The literature on qualitative research has focused on ways researchers make sense of data through coding, memoing and writing. The assumption is that the data are there. But the experience of theory construction is more often one of building jigsaws from pieces of data
and observation in studies where data take many forms and the researcher's interpretation is itself data. In that setting, our experience is that quite normally the jigsaw will have missing pieces, whose "missingness" is the key to understanding.

This paper resulted from discussions with novice and experienced researchers concerned at how qualitative research literature portrays the relationship between theory and data. It is rare to find research accounts that do not make the emergence of theory appear a smooth, even inevitable process. Our own experiences, and those of our students, have never fitted such smooth images, and in discussions we have often found that others are helped by our accounts of the puzzles and anxieties, and the hard detective work, which we have experienced during the analysis stage when a picture appeared to be emerging, but jigsaw pieces were evidently missing.

Missing data are seldom reported in accounts of the research process. To discover that data do not adequately cover the subjects required is thus often interpreted as a personal failure. The doubts are manifold, revolving around the appropriateness of the research question, the data collection methods, the analysis of data and ultimately the researcher’s ability to do research.

In this paper we argue, on the contrary, that in qualitative research it is inevitable there will be missing data, for two different sets of reasons.

Firstly, missing data is always an issue since the data collection processes are not designed in advance to answer a known question. In a grounded study where the aim is to discover theory from data (Glaser & Strauss, 1967) (Strauss & Corbin, 1990), the theory will always require more data, and the processes of "theoretical sampling" are always only partial. The data are missing even when researchers collect and analyze data concurrently so as to be able to adapt interviews to new questions as they emerge (Miles & Huberman, 1984). A common experience in qualitative work is that the research question evolves as
understanding grows. Thus the question at the end of the study differs from the one that initiated the study. The gaps in the data thus emerge when the initial data are viewed through the lens of this subsequent question. This is not a matter for concern. Such a change in the research question is a normal aspect of research, indicating progress in understanding. Popper (1972) argues that the change in question is an essential characteristic of the growth of knowledge. All knowledge grows by an initial problem generating a tentative theory. Critical discussion leads to a process of error elimination which in turn generates the next problem.

Secondly, since meanings and interpretations are the stuff of qualitative studies, data are always missing because there are some questions that respondents do not want to answer, others they circumvent, and most significantly, aspects of their experiences are so taken for granted that they are not reported. In such situations, the silence on these issues in itself becomes powerful data, and the aware researcher can focus on such silences as input to the processes by which they generate theory. It is the “missingness” of data which often provides the most significant theoretical challenges. The more massive the missingness of data, the more likely it is going to be a major theoretical indicator.

Grounded theory has well-tested ways to deal with a change of question. Researchers collect and analyse data at the same time, so as to keep redirecting the research to emerging themes. Theoretical sampling, where data are collected to test theory and the “validity of findings” (Finch & Mason, 1990), is often useful. The 'missingness' problem arises when new questions come up so late in the study that further collection or recoding of data is not possible. The fit of data and theory (Glaser, 1978), an essential part of grounded theory, is often found wanting during the final writing up stage of research. The scope for recoding or re-collecting data is particularly limited in funded studies where external guidelines have to
be met. New questions have then to become the starting point of subsequent research projects.

Once we accept that in the normal course of research there will be missing data, then we need to ask: In the absence of data, how are core categories and emerging theories tested? These are important methodological issues, which are seldom articulated in the literature.

In this paper we examine two kinds of missing data – that is data that are missing because of a change in the question towards the end of a study and emerging theory associated with silence about emerging themes. We examine these through stories about two unrelated areas of study – the social shaping of money and the loneliness of suburbia. In Supriya Singh’s (1997) study of *Marriage Money*, essential data were missing because the initial question had assumed that money was wholly a market phenomenon. The emerging theory however pointed to the existence of different kinds of money in marriage and the market. Questions about power and equality were circumvented by talking about jointness and sharing in marriage money. This in turn led to the core concept of “ritual information” and the essential characteristics of marriage money. In Lyn Richards’ (1990) study of suburbia *Nobody’s Home*, the dominant themes of loneliness and class were not spoken about. Asking why they were not talked about, led to the theory that the interlocking ideologies of family, home ownership and equality had such resilience that they blocked the expression of loneliness and class.

The process of dealing with the “missingness” of data was different, reflecting the different conduct of the studies. With *Marriage Money* the theory was tested by reanalysing the data and testing it against existing studies of marriage and money. The central theme of the social shaping of money was also explored in different contexts through subsequent studies of the use of electronic money in the home and in small business. The issue of missing
data however remained constant as the research questions continued to change with each study.

*Nobody’s Home* was a five year longitudinal ethnographic study with multiple researchers. It had built into it at least three data collection points. These were not theory driven, but each stage of data collection then became a form of implicit or explicit theoretical sampling.

**Marriage Money**

The central theme of the study of *Marriage Money* (Singh, 1997) was that there are different kinds of money in marriage and banking. *Marriage money* and *banking money* differ from the ideal construct of *market money* which underlies economic theory, policy and law. The study of marriage and banking money also led to the conclusion that that there is an intimate interrelationship between the economic and non-economic aspects of social and cultural life.

**Missing data**

The study of marriage money was initially fashioned to answer the question: *How has deregulation changed the relationship of banks and consumers in Australia since 1983?* Two sample sets were chosen to probe how banking patterns differed according to income, education and ethnicity. Singh conducted open ended interviews with 37 persons from 21 middle-income Anglo-Celtic households in a Melbourne suburb. Midway through the interviews, Singh surveyed 188 persons of non-English speaking background with low income and difficulties with English literacy. The data were collected between June 1991 and February 1992. It was only in May 1993, when Singh began writing up her thesis, that “money” emerged as a core category. The ideas relating to “marriage money” and “banking money” were honed in the writing process and did not appear at all in the coding index.

Looking back at the early interviews, it was clear that when Singh asked about changes in banking, the couples talked about the way they handled money in their marriage. This led to
additional interview questions which probed attitudes to money in their marriage, without questioning the initial thrust of the study. Prior theory which came very much from banking literature; and a personal hesitation to probe money in Anglo-Celtic marriage contributed to continuing the initial emphasis on banking. Theoretical sampling was done at this stage to discover whether there were socio-demographic differences in banking patterns. A survey of non-English speaking people with literacy problems was conducted. But the survey had little to say about the nature of marriage money and banking money, for it was nearly wholly focused on banking patterns. This meant that there were large chunks of missing data because of a late change in the research question. The interview data were richer for people talked of money in marriage, despite the fact that Singh did not directly ask them about it. However there was little data on how they constructed the meaning of marriage money, and how these meanings had changed. As the theoretical sampling had not addressed these questions, there was also no opportunity to test whether the meanings of marriage money differed according to marital status, age, income, ethnicity and education.

The data on money and gender were also limited. The data referred to men and women’s interpretation of sharing and commitment in marriage, rather than directly addressing the different ways men and women related to marriage money, banking money, and market money.

There was another kind of missing data because people circumvented troubling questions. Singh had asked about power and equality in the context of money in marriage. However, analysing the data with the help of NUD*IST, a computer programme for the analysis of qualitative data\(^1\), revealed that only three of the 37 persons interviewed mentioned equality and that too, only once. The concept of power came up in two cases, except where Singh introduced

\(^1\) A more detailed analysis of the role of NUD*IST in the analysis of qualitative data can be found in Singh 1996.
The silence about power and equality was so potent, that these concepts disappeared from the coding. This gap in the data about power, equality and money was so profound, that it made Singh question her abilities as a researcher and wonder whether the silence represented a personal blockage.

The realisation that there were missing data on marriage money, power and equality led Singh back to the existing data, and to search for a new theoretical framework. The literature on banking, deregulation and competition which had generated the initial question, could not help explain the nature of multiple monies and people’s silence about power and equality. This search led Singh to investigate the sociology of money; anthropological discussions of meaning and secular ritual; and constructs of information in information science. The new theoretical underpinnings were used to reinvestigate whether the data fit the emerging theory of the social and cultural shaping of money in marriage and banking. The use of NUD*IST made it possible to go back to the data and search it with new themes in mind.

This reanalysis also led to the development of the concept of ritual information. This concept refers to the ritual channelling of information so that some meanings are stressed, while others are blocked. The joint bank account served as a secular ritual which blocked the discussion of power and equality, while stressing aspects of jointness and sharing. This construction was powerfully vindicated by the exceptional cases where equality and power had been mentioned.

Testing the theory in subsequent projects

The theory of multiple monies and ritual information emerged too late in the study to seek more data to test the theory. The social shaping of money was tested and refined through three subsequent projects – The Use of Electronic Money in the Home (Singh, 1996), Trust and Electronic Money (Singh and Slegers, 1997) and Small Business and Electronic Commerce (Singh and Slegers, 1998). The first project took as a given the conceptual framework of money
as a social and cultural phenomenon. The initial research question probed how income, education and gender influenced the access and use of electronic money in the home.

The data substantiated the theory of multiple monies, but not all the data could be explained in terms of socio-economic factors which influenced access and technical literacy relating to the use of electronic money. One case study in particular stood out, where three people in the same high income household used electronic money differently despite having the same degree of access and sufficient expertise in the use of PCs and online technologies. Gender and age alone did not explain the difference. The essential difference was that the woman who did not use the Automated Teller Machines (ATMs) and Electronic Funds Transfer at Point of Sale (EFTPOS) did not trust electronic money. She continued to withdraw cash from the branch once a fortnight, saying, “I wouldn't be sure the bank was really going to transfer the money just if I was pressing a few buttons on a wall”.

This case study alerted Singh to the importance of trust in the use of electronic money in particular, and the perception of money in general. However, as with the *Marriage Money* study, “trust” had not been central to the questions asked in the interviews or in the initial coding process. It was missing data in both senses – missing because the question had emerged late in the study, and missing in the sense that people do not easily talk of “trust”.

Going back to reanalyse the data, Singh and Slegers discovered that people did not speak about forms of money they trusted. They did not dwell on the fact that they trusted cash and cheques. They spoke of trust only when there was a lack of trust. Hence they spoke of trust and money most clearly in contexts where the risk of loss was great and where information was less easily available: in depositing rather than withdrawing; paying large sums as against small sums; in situations where non-payment was going to result in loss or penalty; paying with new forms of electronic money; payment over open rather than closed networks.
Trust is difficult to define because it is nebulous and all pervading. People speak of trust in terms of the factors that go to engender trust, such as comfort, certainty, being more confident and in control, faith and reliability. They also speak of trust in terms of security and related aspects such as fraud, theft and privacy.

It was also realised that trust had not been an initial focus of the study of the use of electronic money because “security” rather than “trust” is at the centre of the discussion of electronic commerce in policy and industry circles. The emerging theme of trust led Singh and Slegers to the multidisciplinary literature on trust; to distinguish between trust and security; and to delineate how trust is engendered in an online environment. The new theoretical readings, together with the data, led Singh and Slegers to postulate that trust in electronic money is engendered by giving the customer control, comfort and the perception of caring.

This theory in turn was further tested in the context of small business, where it was found that the lack of trust made payments the last activity to go online. It also led to the finding that when communication was wholly online, information had to be more explicit, because of the absence of non-verbal communication and the lack of the physical and cultural context of information. The explicitness of information at times led to a redefinition of the activity. With payments this was particularly relevant, for the emphasis online moved from payment to the broader activity area of money management. Again this new theory meant that there was missing data which had to be explained, a new theoretical framework sought, existing data re-examined, and the theory tested with further data.

This sequence of research substantiates that the research process constantly has to deal with missing data, as it was the gap between data and theory that triggered the emergence of new theory and research problems. Fitting data to theory is the process that reveals the gaps. This
process modifies theory. It helps interpret the silence, but the questions that emerge from this interpretation, leave new questions and hence incomplete data.

**Nobody’s Home?**

The five-year Green Views Project was designed as an ethnographic study of outer suburbia, initiated on methodological grounds. The literature on suburban isolation, with few exceptions, was static, one-off pictures of existing suburbs. The methodology, again with few exceptions, was survey research. Richards’ earlier studies of Australian families had convinced her that one-off surveys were of limited use in understanding the processes of family construction, the social contexts of reported experience or the ways people met and managed their relationships with ideology. In particular, women’s accounts of their isolation and loneliness during the years of young motherhood were to be set in context.

The design of the Green Views project was explicitly different. The project team would be involved in an outer suburban estate from its first planning stages, following through not only the arriving residents but also the life of support networks, resident groups and issues. The goal was to see processes, to contextualise the accounts of family life, and to try to understand traditional family ideology in action, in the hard core of the traditional family. There was no formal initial research question, but certainly one persistent question was “How do the residents in a new outer suburb cope with isolation and loneliness?”

*“Your fault if you’re lonely”*

At the end of the first year, the body of survey and qualitative data was formidable, and the build-up of data had had several results. One was the early work of Lyn and Tom Richards towards computer programming of such data in the NUD*IST project; the resultant software supported (and its development inevitably delayed!) analysis for the remainder of the project. Using NUD*IST, Richards became increasingly confident of her ability to access the rich
participant observation and interview data. (Richards, L. and Richards, T. 1994; Richards, T. and Richards, L., 1994; Richards, 1996) But that confidence brought her up against the first of five areas of missing data. The better the access to data, the clearer it became that none of those talking to the researchers were reporting that they were lonely

----------------------------
Insert figure 1 about here
----------------------------

Was nobody lonely in this suburb where the warning, “You won’t be able to do interviews here, nobody’s home!” had already named the future book? A first survey perhaps had been too soon for loneliness to have set in? Certainly it offered cheerful statistics on number of neighbours “known” (but what did people mean by “know”?) Careful analysis of the data on neighbouring did indeed establish an early period of social activity, but also that it was usually superficial and transient. Researchers reported a vocabulary of descriptors for this “falling-off” period: relationships with the neighbours “go off” or “go sour” after a while (like cream). Yet in all our open-ended interviews and observations, we were gathering none of the classic accounts of isolation and loneliness. On the contrary, in this elegant upwardly mobile “family community”, several people told us, “It’s your fault if you’re lonely.” The theory underpinning the study, about social isolation in outer suburbs, and about home ownership and the costs of privacy and loss of community, seemed severely challenged. Only a flat rejection of the data, and assertion that the informants were lying or suffering from false consciousness, would make those theories fit the data. Richards took into the next stages of the study the concern that the data appeared to force such interpretation.
“The migrants don’t matter”

The study proceeded towards a second year and a reinterview, with other areas of data, as in all ethnographies, overtaking the team’s attention. Most pressing were issues of ethnic diversity. Demographically, it was obvious: very high proportions of residents had come from the inner western suburban areas, and were of first or second generation migrant origin, the largest group being Greek. Other residents who had moved “out” to get away from those areas would clearly be highly likely to resent that they had been followed by the migrants they were moving from.

Missing data again. Nobody said so, in surveys or casual conversation. The team were told the Greek children at the school were being beaten up and called wogs, but the parents of those same children spoke of the pleasures of ethnic diversity. With the follow-up survey ahead, Richards inserted direct questions about attitudes to the migrants, and a “wogs go home” slogan on the green fence surrounding the estate. The most common answer was that “the migrants don’t matter” (like flies?), and the most common negative response to the graffiti was anger that vandals would do that to our fence, making the estate look bad. (Only one woman broke the picture of homogeneous acceptance: she said seriously that she thought her husband might have written it.)

The theme might have been left had a very detailed case study not been presented (sheer good luck allowing the combination of rich interview material from both sides of a very distressing conflict in one of these quiet courts). The data clearly established the overt prejudice to Greek neighbours by a couple well known to and liked by the whole research team - and juxtaposed the data from their earlier interviews with the clear evidence of its contradiction.

“All in the same boat”

Meanwhile, an obviously related third body of missing data had appeared, and Richards was for the first time thinking of missing data in those terms. The third was about class. The
researchers rapidly discovered in this most obviously upwardly mobile estate that the language of class did not appear.

This body of missing data, unlike the earlier ones about loneliness and prejudice, was directly challenged. Indeed probing to “get” informants talking about class became a much-discussed challenge. It took some time to discover that whilst we watched, the residents were creating a microcosm of social differentiation within the suburb, with finely drawn social areas (“the hill,” “down by the creek” etc.) almost immediately mirrored in a wide range of land prices and behaviour in the action groups.

“A terrible lot of lamingtons”

The groups were meanwhile becoming a major focus. Group life (or lack of it) had been studied from the earliest activity, and the activity had become increasingly acrimonious. A new puzzle; in this most harmonious-imaged family community, why were the groups so contentious? Why did conflicts so divide residents, and issues that seemed trivial (the prize for a raffle, the location of a meeting) cause such long-drawn-out conflicts?

Missing data. There simply were no explanations for the divisive conflicts that made sense in a political analysis of issues and power. The offered explanations were bizarre, and the observations rapidly produced a rich body of data that fascinated and distracted. (Miles’ account
of qualitative data as “an attractive nuisance” (1979) fitted wonderfully this stage in the project!
In one group, a blow-up resulted in one couple selling their house and leaving the estate, others resigning from the group. One woman explained the behaviour of another by reminding the researcher she “had sold a terrible lot of lamingtons”.

The episode led Richards into what she at that stage saw as a quite unconnected area, the sociology of fundraising - and thus into (or back to) the sociology of ideology - gender, volunteering, unpaid labour. And thus, back to the earlier missing data. What was it about this most un-communal “community” that obscured loneliness, prejudice and class, in a smudging apparent homogeneity, yet allowed trivial issues to make massive conflicts?

“There when you need them but not in your pocket”

Richards used the second survey to stalk the missing data on community and loneliness. Rather than the standard and apparently useless questions about how many people you “know” and how many groups you “belong” to, she tried an indirect approach. Perhaps you could see these hinted but invisible bodies of missing data by a sort of methodological equivalent of averted vision. (Focus a little to the side of a celestial faint object and you see it more clearly.) The survey focussed close - seeking, with a quirky questionnaire, portraits of the sort of relationships people had with those they know best. Then the focus zoomed out, with questions seeking “off-the-hook,” easily accessed definitions that might indicate background, taken for granted dominant ideology.

Cluster analysis showed six very robust different types of “best known” neighbour relations, the types strongly associated with demographic dividers, especially age and family stage. By contrast, the general questions showed very common, strong views on good neighbours, not divided by gender or age, with a common off-the-hook vocabulary (“too close for comfort,” “not in your pocket”).
**Missing Data Analysis**

In each of the areas sketched above, the missing data was an identifiable hole in the project, not merely something not asked about. Like Singh, Richards arrived at the linking of these themes late in the analysis (Richardson, 2000). The outcome of that linking is in the book (1990), and like Singh, she found that much of the cohering happened during writing. First one, then another of the areas of missing data slid into place in an account of the twinned ideologies of family and home ownership and the residents’ attempts to negotiate with those. Interlocking ideologies acquire a resilience, it was concluded, making challenges not merely harder but literally unthinkable, since they are not thought about.

In each area, like Singh, Richards was helped by having detailed and fine-fingered access to data through NUD*IST, and the ability to keep an audit trail of understanding. Both projects in turn contributed to the further development of the software! (Richards, 1997). And the experiences of stalking missing data informed the design of a next project (Richards, 1995, Richards, Seibold and Davis, 1997).

In each area, the missing data started to emerge from behind ideology when experience and social process was seen from the sidelines, through detailed knowledge of events or cases, through indirect questions, or pursuit of patterns in vaguely related attitude statements.

**Missing Data and Theory**

The experience of Singh and Richards suggests that missing data are important both for the emergence and testing of theory and for their link to the discovery of new questions.

It is often in testing previous theory against the data that it is realised that some data are missing, in the sense that people have chosen not to speak of the questions they were asked. . . Missing data are particularly important for understanding the way people construct meaning in different social and cultural contexts. For Singh and Richards, it was the realisation there were
important gaps in their respondents’ narratives, that led them to understand how people had constructed the meanings of money, family and suburban life.

Missing data also lead to new research questions. These questions in turn reveal the missingness of data, leading to reanalysis of existing data and the search for new data to test the emerging theory. Missing data are thus not only an inevitable part of the research process, but also represent one of the most challenging stages for the emergence of theory and the discovery of questions. It is our hope that in recounting our experiences of these research conundrums we can assist others in presenting their journey to explanation, including dead ends and lost directions, as well as satisfying arrivals.

References


Figure 1: Suburban Loneliness: The Initial Question, Data and Theory

Q. How do the residents in a new outer suburb cope with isolation and loneliness?

Data: Field research, interviewing of new arrivals in Green Views

“Your fault if you’re lonely”

Theory: Social isolation in outer urban areas
Home ownership ideology and its costs

Theory: Loneliness is a secret - why?
Figure 2: Apparent classlessness in the symbolic centre of class-ridden society

Q. What are the class origins and class consciousness of these newly successful residents of outer suburbia?

Data: Field research, interviewing direct and indirect probing

“All in the same boat”.

NOBODY talks of class (except the kids)

The people on the hill - social construction of difference

Asking a different question; the local creation of classes…
Q. How do groups form and when do they work for residents?

- Data: participant observation and interviewing in major groups in Green Views over 3 years

Q. What are the sources of the failures of groups and many conflicts?

- "a terrible lot of lamingtons!"
- the big issues are all very trivial!
- No data worthy of political analysis?

Literature on ideologies of gender, class, volunteering

The social construction of homogeneity

Testing it with new data