

How to do (or not to do) . . .

Domain analysis for qualitative public health data

SARAH ATKINSON¹ AND MÔNICA ABU EL HAJ²

¹*Geography Department, University of Manchester, UK, and* ²*Escola de Saúde Pública, Ceará, Brazil*

Introduction

Many researchers in public health are drawing on multiple methods to collect data in order to answer applied questions in the field. In particular, the value of qualitative data is recognized more and more as a supplement to existing or primary quantitative data which can serve either to refine research questions or to assist in the interpretation of statistical results. Manuals are plentiful on how to collect qualitative data relatively quickly through group discussions or interviews with key actors (for example Krueger 1988; Scrimshaw and Hurtado 1987; WHO ARI programme 1993). However, less advice seems to be readily accessible on what to do with those data once collected.

The steady stream of research students seeking advice that we have experienced knocking on our doors as social scientists based in public health schools, together with a tendency towards apology in presentation of qualitative data in multi-method research theses and the often somewhat dismissive attitude towards qualitative data by public health managers, have highlighted the need for accessible guides on how to analyze the data. This short piece introduces one easy approach for analyzing the content of qualitative data. The approach draws on that of Spradley (1979) and is based on the identification within the content of the data of key topics, referred to as domains, and the relationships between them.

Two general points about this approach need making at the outset. First, the approach is focused entirely on the content of verbal or written communications and is intended for those researchers who are using qualitative data as one amongst a number of methods within public health. Most commonly, these data will be serving as a supplement to help interpret data from

other sources, rather than being the primary data source in the research. Researchers coming from more of a social science perspective who want to analyze language use in greater depth will need to go well beyond the approach suggested here and explore texts on discourse analysis such as those of Burman and Parker (1993) or Fairclough (1995). Secondly, this introduction to the analysis of qualitative data needs to be used in a context of wider reading and training in the theories and methods of the social sciences and not as a quick fix guide to be used alone.

The method is presented here with reference to open interviews, but the approach is applicable to any source of data that is unstructured and aiming to allow the participants to identify for themselves the topics and issues of importance. Examples to illustrate the processes are presented from interviews held with pregnant women in Northeast Brazil on their views of the local antenatal services (Atkinson 1993).

Interviews can be analyzed singly, emphasizing individual continuity and coherence, or together representing an identifiable social group with a common overview on the research question concerned. Generally, for qualitative data within the public health field, the research question is likely to require a primary focus on the common categorizations and perceptions of specified social groups, and thus it is assumed that data from a number of interviews or group discussions are being analyzed together (see Weatherall and Potter 1988 for a discussion of individual versus collective analyses). The analysis is made through four steps, adapted from Spradley (1979) and we have retained his rather heavy terminology for each of the steps: the identification of the main issues or 'domains' raised by the interviewees;

the grouping of more detailed topics within each of these domains so as to construct a 'taxonomy' of sub-categories; the specification of what was actually said, the 'components' within each sub-category; and finally the construction of an overall picture by exploring the inter-relationships between the various domains.

Step 1: Identifying the domains

The first step requires the researcher to identify the primary domains which recur in the interviewees' discourse. The challenge here for the researcher is to ensure that the domains that are defined do reflect the concerns of the interviewees as indicated in their narratives, rather than merely reflecting the researcher's own pre-defined set of categories. Of course, ultimately the domains are defined by the researcher and therefore they do represent an imposition by the researcher of her/his way of classifying topics onto the narrative data. Furthermore, there may be several different and equally valid ways that the information can be categorized.

Nonetheless, there are simple strategies for double-checking the importance to the interviewees of the primary domains identified. First, in the context of public health, the researcher will be concerned primarily with practical matters and thus the categories can be defined in terms of concrete issues raised by the interviewees rather than the linguistic content and so forth which may be of concern to social scientists. Thus the orientation of the domains is already determined. In order to identify those issues, the researcher needs to familiarize herself with the data. The texts should be read through several times. Then each interview can be indexed, recording the topics of discussion line by line. The indexing of the topics which can then be collated across all interviews permits the identification of a preliminary list of issues (Table 1). Some of these will recur more frequently than others and some can be defined as sub-topics of broader categories. Thus, by grouping them together, a final list of just a few fairly broad domains can be made (Table 2).

Step 2: Constructing a taxonomy of sub-categories

Once the primary domains are established, other topics can be defined as secondary sub-sets of them. Rather than simply relying on the preliminary topics

identified through the indexing system, at this stage it is useful to start arranging the actual text into the primary domains. This strategy groups actual phrases together and allows the identification of the sub-categories to emerge directly from the interviewees' own words and thus is more likely to represent those topics most important to the interviewees. An example of the results of this kind of taxonomic analysis is given in Figure 1.

Table 1. Example of a preliminary list of topics discussed

Getting and being pregnant:	Signs of pregnancy, danger signs, planned, physical problems (varicose veins, swollen feet/legs, sleeping, aches)
Feelings during pregnancy:	Nerves, anger/fright, worries, embarrassment, inconvenience, impressions
Aspects of blood during pregnancy:	Pressure, anaemia, weak/strong, thick/thin, iron sulphate pills, vitamins, vaccines/tetanus
Family planning:	Different methods including sterilization
Advice/activities to promote health:	Exercise, activities, smoking, self-care, imprinting, advice sources, information sources
Birth and miscarriage:	Previous experiences, place, signs, caesarean/normal, spontaneous and induced abortion, wet/dry birth, adoption, birth weight, sterilization and birth
Aspects of general lifestyle:	Geography, family, husband(s), children, raising children, paid employment, child care, crèches, housing, residential area, sources of support
Going for antenatal care:	Health staff, place, previous experiences, what happens, meetings, use, tests, distance/cost, logistics, waiting time, discontinued attendance, Saturdays, ultrasound, sterilization, high risk, information give and given, time off work, charges

Source: unpublished field notes, Atkinson, 1992

Table 2. Example of the primary domains identified

Motivations for antenatal care
Medical process [experience of antenatal care and evaluation]
Risks during pregnancy
Reproductive histories
Social and economic background

Source: Atkinson 1993

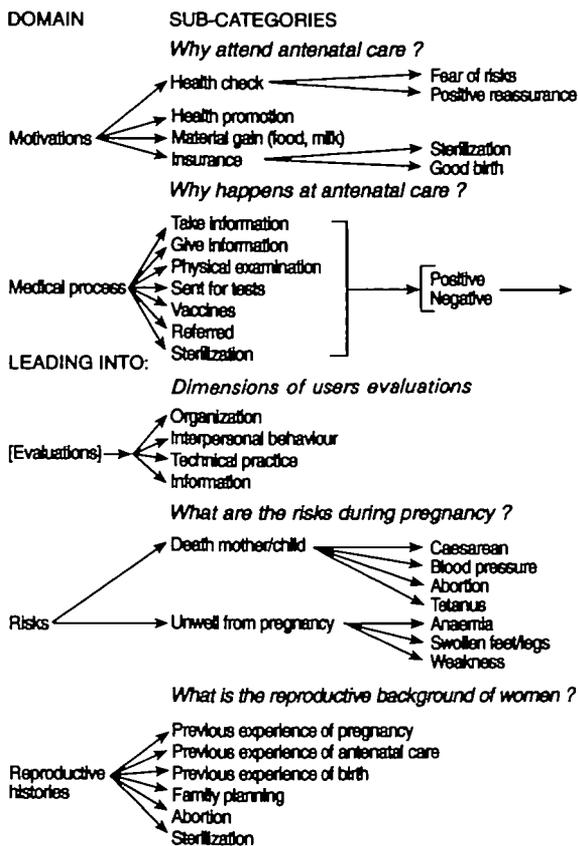


Figure 1. A taxonomic analysis

Step 3: Specifying the components

The third step allows the researcher to finally get to grips with what she/he has really been looking for throughout all of this, namely, what those interviewed actually say about the various sub-categories of topics that have been identified in the first two stages. This is the crucial part of the content for defining where problems or successes lie in the public health aspect under study. By collating all phrases or whole narratives on various sub-categories in the second step, the bulk of the work for this stage has already been done. All that remains is to summarize the content into key issues for planners or policy-makers to address.

In written reports, the results from this step are typically presented as direct quotations from the interviewees. In our study of antenatal care in Northeast Brazil, within a domain of motivations for attending the service and a sub-category of 'health

check', women identified the avoidance of risks as a key factor. This can be reported through their own words, for example:

'you should attend antenatal care for yourself; some women die in pregnancy'

(From unpublished interview transcript, Abu El Haj, 1992)

Step 4: Relating the domains

The last stage in this approach is to identify relations both between sub-categories, if this is appropriate within the research question, and, more importantly, between the primary domains. In this way the researcher tries to build up an overall picture. Within the collection of actual quotations made by interviewees, the researcher needs to seek statements that relate one sub-category to another, either in terms of influence or priority. This is best illustrated by example.

In our study, which aimed to explore user views of health service quality, we adopted two strategies for identifying priorities and influences between the domains and sub-categories. First, we labelled health facilities as to whether women for the most part spoke well or badly of them. Secondly, we looked in more detail at who were the women who were most and least satisfied with the health care they had received. These two procedures lead to the conclusion that there were certain issues which regularly resulted in particularly high dissatisfaction or satisfaction. This, in turn, enabled us to claim that the priorities between issues liked or disliked in health care delivery cut across all sub-categories and thus attention was needed on all aspects of health care. Finally, we were able to establish associations and influence between the primary domains which highlighted the importance of the context of health service users' lives beyond the specific moment of the medical encounter (Figure 2).

Main problems

We have already mentioned the problem of double-checking that the categories the researcher imposes onto the qualitative data do reflect topics given importance by interviewees. This is particularly true where data are collected over a period of time and the researchers inevitably start to identify themes and topics as they go along; indeed this iterative process is one of the most valuable advantages of a qualitative approach. The strategy outlined above of formally

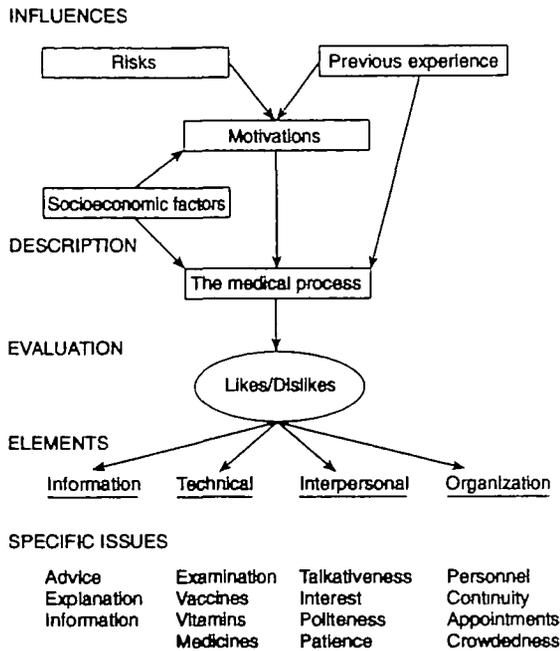


Figure 2. Relationships of the five domains

indexing the transcripts of the interviews can help bring out serious discrepancies between the weight given to a topic by the researcher and by the interviewees as a whole group. This emphasis on treating the interviewees as a whole group relates to a second problem needing consideration.

Amongst the interviews held, researchers will find considerable variation between interviewees or discussion groups in terms of the extent to which lay explanations are expressed. Within the anthropological literature, the researcher is always hoping to meet the so-called key informant, a local informant who has a particular gift for analyzing the everyday experiences and interactions of the population group under study. The problem in a relatively short piece of qualitative fieldwork is how much importance to give to the explanations or information provided by such local analysts.

In the antenatal care study, out of a total of fifty-one women interviewed, we met one woman who developed all kinds of explanations for us regarding risks during pregnancy, the role of blood in the body and so forth. One other woman touched on some of these themes also but to a lesser extent. Our problem then was whether to base an analysis of risk and the

role of blood around the explanations offered by our key informant or whether to blend her information in as just one other interview. Needless to say, we managed to find a way of doing a bit of both. During the data collection, we drew on ideas presented by the key informant to prompt conversation with other women. However, by the end of all interviews, we concluded that her explanations had not been echoed sufficiently strongly by other women, even when prompted directly, to conclude that this formed the basis of a local explanatory model regarding risks during pregnancy. Therefore, we limited the emphasis given to her explanations in our final analysis far more than we had imagined we would do immediately following the interview.

Concluding comments

There is an increasing interest in making use of qualitative sources of data within public health research. However, researchers often feel rather nervous in the face of all the data they have managed to collect once the time for analysis arrives and tend to limit their analysis to simple descriptive reports. On the other side, health managers are also more familiar with quantitative methods and can disregard the implications of research that has a qualitative base. The approach presented here is offered as one way in which public health researchers can analyze and present their qualitative data in a way that stresses the characteristics of systematicness and inter-relationships with which we hope both researchers and the users of their results may feel more comfortable.

References

- Atkinson SJ. 1993. Anthropology in research on the quality of health services. *Cadernos de Saúde Pública* 9(3): 283–99.
- Burman E, Parker I (eds). 1993. *Discourse analytic research*. London: Routledge.
- Fairclough N. 1995. *Critical discourse analysis: the critical study of language*. Harlow: Longman.
- Krueger RA. 1988. *Focus groups: a practical guide for applied research*. Newbury Park: Sage.
- Scrimshaw SCM, Hurtado E. 1987. *Rapid Assessment Procedures for nutrition and primary health care*. California: UCLA; Tokyo: UNU.
- Spradley J. 1979. *The ethnographic interview*. Orlando: Holt, Rinehart, Winston.
- Wetherall M, Potter J. 1988. Discourse analysis and the identification of interpretative repertoires. In: Antaklis C (ed). *Analysing everyday explanations*. Chapter 12: 168–83. London: Sage.
- WHO, ARI programme. 1993. *Focused ethnographic study of acute respiratory infections*. WHO, Geneva.

Biographies

Sarah Atkinson is an anthropologist working on social and organizational aspects of changing health care systems, particularly in urban areas. Her research focuses on the different views that different actors in health systems have of the same policy or intervention and how this may shape the effect of that intervention. She is based in the Geography Department of the University of Manchester.

Mônica Abu El Haj is a freelance anthropologist working in the state of Ceará in the northeast of Brazil, specializing in public policy. Her research experience has included medical anthropological topics such as the diseases of childhood and HIV/AIDS, user perspectives on health service quality, implementation of policy change in the health and education sectors.

Correspondence: Dr Sarah Atkinson, Geography Department, Mansfield Cooper Building, University of Manchester, Oxford Rd, Manchester, M13 9PL, UK.