

A general inductive approach for qualitative data analysis

David R. Thomas
School of Population Health
University of Auckland, New Zealand

Phone +64-9-3737599 Ext 85657
email dr.thomas@auckland.ac.nz

August 2003

An outline of a general inductive approach for qualitative data analysis is described and details provided about the assumptions and procedures used. The purposes for using an inductive approach are to (1) to condense extensive and varied raw text data into a brief, summary format; (2) to establish clear links between the research objectives and the summary findings derived from the raw data and (3) to develop of model or theory about the underlying structure of experiences or processes which are evident in the raw data. The inductive approach reflects frequently reported patterns used in qualitative data analysis. Most inductive studies report a model that has between three and eight main categories in the findings. The general inductive approach provides a convenient and efficient way of analysing qualitative data for many research purposes. The outcomes of analysis may be indistinguishable from those derived from a grounded theory approach. Many researchers are likely to find using a general inductive approach more straightforward than some of the other traditional approaches to qualitative data analysis.

A general inductive approach for qualitative data analysis

There is a wide range of literature that documents the underlying assumptions and procedures associated with analysing qualitative data. Many of these are associated with specific approaches or traditions such as grounded theory (Strauss & Corbin, 1990), phenomenology (e.g., van Manen, 1990), discourse analysis (e.g., Potter & Wetherall, 1994) and narrative analysis (e.g., Leiblich, 1998). However some analytic approaches are “generic” and are not labelled within one of the specific traditions of qualitative research (e.g., Ezzy, 2002; Pope, Ziebland, & Mays, 2000; Silverman, 2000). In working with researchers who adopt what has been described as a “critical realist” epistemology (Miles & Huberman, 1994) the author has found that many researchers unfamiliar with any of the traditional approaches to qualitative analysis, wish to have a straightforward set of procedures to follow without having to learn the technical language or “jargon” associated with many of the traditional approaches. Often they find existing literature on qualitative data analysis too technical to understand and use. The present paper has evolved from the need to provide researchers analysing qualitative data with a brief, non-technical set of data analysis procedures.

A considerable number of authors reporting analyses of qualitative data in journal articles (where space for methodological detail is often restricted) describe a strategy that can be labelled as a “general inductive approach.” This strategy is evident in much qualitative data analysis (Bryman & Burgess, 1994; Dey, 1993), often without an explicit label being given to the analysis strategy. The purpose of the present paper is to describe the key features evident in the general inductive approach and outline a set of procedures that can be used for the analysis of qualitative data. The inductive approach is a systematic procedure for analysing qualitative data where the analysis is guided by specific objectives.

The primary purpose of the inductive approach is to allow research findings to emerge from the frequent, dominant or significant themes inherent in raw data, without the restraints imposed by structured methodologies. Key themes are often obscured, reframed or left invisible because of the preconceptions in the data collection and data analysis procedures imposed by deductive data analysis such as those used in experimental and hypothesis testing research.

The following are some of the purposes underlying the development of the general inductive approach. These purposes are similar to other qualitative analysis approaches.

1. To condense extensive and varied raw text data into a brief, summary format.
2. To establish clear links between the research objectives and the summary findings derived from the raw data and to ensure these links are both transparent (able to be demonstrated to others) and defensible (justifiable given the objectives of the research).
3. To develop of model or theory about the underlying structure of experiences or processes which are evident in the text (raw data).

The general inductive approach is frequently reported in health and social science research. The following examples of descriptions, taken from the methods sections of

research reports, illustrate data analysis strategies that have used a general inductive approach.

The transcripts were read several times to identify themes and categories ... In particular, all the transcripts were read by AJ and a subsample was read by JO. After discussion a coding frame was developed and the transcripts coded by AJ. If new codes emerged the coding frame was changed and the transcripts were reread according to the new structure. This process was used to develop categories, which were then conceptualised into broad themes after further discussion. The themes were categorised into three stages: initial impact, conflict, and resolution. (Jain & Ogden, 1999, p. 1597)

Emerging themes (or categories) were developed by studying the transcripts repeatedly and considering possible meanings and how these fitted with developing themes. Diagrams were used to focus on what was emerging and to link patient and doctor themes into major barriers to referral. Transcripts were also read “horizontally”, which involved grouping segments of text by theme. Towards the end of the study no new themes emerged, which suggested that major themes had been identified. (Marshall, 1999, p. 419)

A rigorous and systematic reading and coding of the transcripts allowed major themes to emerge. Segments of interview text were coded enabling an analysis of interview segments on a particular theme, the documentation of relationships between themes and the identification of themes important to participants. Similarities and differences across sub-groups (e.g. service providers vs individuals, recent vs long-term migrants) were also explored. (Elliott & Gillie, 1998, p. 331)

The inductive approach is evident in several types of qualitative data analyses, especially grounded theory (Strauss & Corbin, 1990). It is very similar to the general pattern of qualitative data analysis described by others (e.g., Miles & Huberman, 1994, p. 9; Pope et al, 2000). Inductive approaches are intended to aid an understanding of meaning in complex data through the development of summary themes or categories from the raw data (“data reduction”). These approaches are evident in many qualitative data analyses. Some have described their approach explicitly as “inductive” (e.g., Backett & Davison, 1995; Stolee, Zaza, Pedlar, & Myers, 1999) while others use the approach without giving it an explicit label (e.g., Jain & Ogden, 1999; Marshall, 1999)

Underlying assumptions

Some of the assumptions, which can be seen as underlying the use of a general inductive approach, are described below.

1. Data analysis is determined by both the research objectives (deductive) and multiple readings and interpretations of the raw data (inductive). Thus the findings are derived from both the research objectives outlined by the researcher(s) and findings arising directly from the analysis of the raw data.
2. The primary mode of analysis is the development of categories from the raw data into a model or framework that captures key themes and processes judged to be important by the researcher.

3. The research findings result from multiple interpretations made from the raw data by the researchers who code the data. Inevitably, the findings are shaped by the assumptions and experiences of the researchers conducting the research and carrying out the data analyses. In order for the findings to be usable, the researcher (data analyst) must make decisions about what is more important and less important in the data.
4. Different researchers are likely to produce findings which are not identical and which have non-overlapping components.
5. The trustworthiness of findings can be assessed by a range of techniques such as (a) independent replication of the research, (b) comparison with findings from previous research, (c) triangulation within a project, (d) feedback from participants in the research, and (e) feedback from users of the research findings.

Features of categories developed from coding

The outcome from an inductive analysis is the development of categories into a model or framework that summarises the raw data and conveys key themes and processes. The categories resulting from the coding, which are the core of inductive analysis, potentially have five key features.

1. Label for category: Word or short phrase used to refer to category. The label often carries inherent meanings that may not reflect the specific features of the category.

2. Description of category: Description of the meaning of category including key characteristics, scope and limitations.

3. Text or data associated with category: Examples of text coded into category that illustrate meanings, associations and perspectives associated with the category

4. Links: Each category may have links or relationships with other categories. In a hierarchical category system (e.g., tree diagram) these links may indicate superordinate, parallel and subordinate categories (e.g., *parent*, *sibling* or *child* relationships). Links are likely to be based on commonalities in meanings between categories or assumed causal relationships.

5. Type of model in which category is embedded: The category system may be incorporated in a model, theory or framework. Such frameworks include; an open network (no hierarchy or sequence), a temporal sequence (e.g., movement or time), or a causal network (one category causes changes in another). It is also possible that a category may not be embedded in any model or framework.

The process of inductive coding

Inductive coding begins with close readings of text and consideration of the multiple meanings that are inherent in the text. The researcher then identifies text segments that contain meaning units, and creates a label for a new category into which the text segment is assigned. Additional text segments are added to the category where they are relevant. At some stage the researcher may develop an initial description of meaning of

category and by the writing of a memo about the category (e.g., associations, links and implications). The category may also be linked to other categories in various relationships such as: a network, a hierarchy of categories or a causal sequence.

The following procedures are used for inductive analysis of qualitative data.

1. Preparation of raw data files ("data cleaning")

Format the raw data files in a common format (e.g., font size, margins, questions or interviewer comments highlighted) if required. Print and/or make a backup of each raw data file (e.g., each interview).

2. Close reading of text

Once text has been prepared, the raw text should be read in detail so the researcher is familiar with the content and gains an understanding of the "themes" and details in the text.

3. Creation of categories

The research identifies and defines categories or themes. The upper level or more general categories are likely to be derived from the research aims. The lower level or specific categories will be derived from multiple readings of the raw data (in vivo coding). For "in vivo" coding, categories are created from meaning units or actual phrases used in specific text segments. Several procedures for creating categories may be used. Copy and paste (e.g., using a word processor) marked text segments into each category. Specialist software (e.g., NUD*IST, Ethnograph) can be used to speed up the coding process where there are large amounts of text data.

4. Overlapping coding and uncoded text

Among the commonly assumed rules that underlie qualitative coding, two are different from the rules typically used in quantitative coding: (a) one segment of text may be coded into more than one category. (b) a considerable amount of the text may not be assigned to any category, as much of the text may not be relevant to the research objectives.

5. Continuing revision and refinement of category system

Within each category, search for subtopics, including contradictory points of view and new insights. Select appropriate quotes that convey the core theme or essence of a category. The categories may be combined or linked under a superordinate category when the meanings are similar.

An overview of the coding process is shown in Table 1. The intended outcome of the process is to create three to eight summary categories, which in the coder's view captures the key aspects of the themes in the raw data and which are assessed to be the most important themes given the research objectives. Inductive coding which finishes up with more than about eight major themes can be seen as incomplete. In this case some of the categories may need combining or the coder has not made the hard decisions about which themes or categories are most important.

Table 1: The coding process in inductive analysis

Initial read through text data	Identify specific segments of information	Label the segments of information to create categories	Reduce overlap and redundancy among the categories	Create a model incorporating most important categories
Many pages of text	Many segments of text	30-40 categories	15-20 categories	3-8 categories

Note: Adapted from Creswell, 2002, Figure 9.4, p. 266

Example of inductive coding

In a study analysing qualitative data on therapy effectiveness for women who had reported experiencing childhood sexual abuse (CSA), Kim McGregor reported on her analysis of responses to the question “What first led you to see a therapist for the effects of childhood sexual abuse?”

Participant B gave the response: “*My partner took me because he felt that the effects of my CSA were affecting our relationship.*” This response was considered to contain two different meaning units and therefore was counted as two units of meaning, and was assigned two different code numbers that related to two different categories. The first text segment of meaning “*my partner took me*” fitted into the category “Referred by family.” The second text unit of meaning was determined to be: “*because he felt that the effects of my CSA were affecting our relationship*” – this text segment was considered to fit into the category “Relationship difficulties.” Thus these two text unit meanings were counted as two units of the total 315 text units that emerged from Qn 5. (McGregor, 2003, p.82)

In this example the author has described how specific text segments were examined for meaning and assigned to the emerging categories. In all she reported 22 categories for responses to the question above. Some of these “lower level” (“open”) categories were subsequently combined.

The 22 categories that emerged ... were later reduced to 15 categories with some of the smaller categories being merged with similar allied categories. For example, three of the 22 different categories that emerged from the text units in Qn 5 were labelled ‘Depression’, ‘Breakdown’ and ‘Suicidality’. However it became difficult to fit some text units neatly into only one of these categories. These three categories appeared to be on a continuum of despair. The boundaries between each of these categories seemed to be blurred, and a more meaningful category (that reflected the meaning from the data) seemed to be a merger of these three categories. The amalgamated category retained the ‘feel’ of the continuum of despair by merging the three labels into one: ‘Depression/Breakdown/Suicidality’. (McGregor, 2003, p.82)

Assessing trustworthiness

Procedures that can be used for assessing the trustworthiness of the data analysis include *consistency checks* (e.g., having another coder take the category descriptions and find the text which belongs in those categories) and credibility or *stakeholder checks*. Stakeholder checks might involve opportunities for people with a specific interest in the research, such as participants, service providers, funding agencies, to comment on categories or the interpretations made (Erlandson, Harris, Skipper & Allen, 1993, p. 142). Comparisons may be made with previous research on the same topic. The usefulness of the research findings for policy and services planning is another possible credibility check. Stakeholder checks are different from triangulation, where data from one source is checked for consistency with data from other sources. Some examples of consistency and stakeholder checks are set out below.

Independent coding: An independent coder is given the research objectives and some of the raw text from which the categories were developed. They are asked to create categories from the raw text.

Coding consistency check: An independent coder is given the research objectives, the categories and descriptions of each category, without the raw text attached. They are then given a sample of the raw text (previously coded by the initial coder) and asked to assign sections of the text to the categories that have been developed. The raw text selected has sections of text from which the initial categories were derived. A variation of this type of check is to give an independent coder both the initial categories and some of the text assigned to these categories. They are then given text that has not been coded and asked to assign sections of the new text into the initial categories.

Stakeholder checks: Stakeholder checks enhance the credibility of findings by allowing research participants and other people who may have a specific interest in the research to comment on or assess the research findings, interpretations, and conclusions. Such checks may be important in establishing credibility for the research findings. For example, participants in the settings studied are given a chance to comment on whether the constructions of the researcher relate to their personal experiences. Stakeholder checks may be carried out on the initial documents (e.g., interview transcriptions and summaries) and on the data interpretations and findings. Stakeholder checking may be conducted progressively during a research project both formally and informally. Listed below are procedures describing possible stakeholder checks during a research project.

- At the completion of interviewing by summarising the data and allowing respondents to immediately correct errors of fact or challenge interpretations
- During subsequent interviews by asking respondents to verify interpretations and data gathered in earlier interviews
- In informal conversations with members of an organization with interests in the setting being studied
- By providing copies of a preliminary version, or specific sections, of the research report to stakeholder groups and asking for written or oral commentary on the report.
- Before submission of the final report, a stakeholder check may be conducted by providing a complete draft copy for review by respondents or other persons in the setting being studied.

Writing the findings

When reporting findings from inductive analysis, the summary or top-level categories are often be used as main headings in the findings, with specific categories as subheadings. It is good practice to include suitable quotes in the text to illustrate the meanings of the categories. Several examples are outlined to indicate how categories are labelled and described, and how authors reporting qualitative findings use detailed descriptions and quotations to illustrate the meanings of the categories developed.

Quality of nursing care

The following description, taken from Williams and Irurita (1998), illustrates a common reporting style for analyses derived from inductive or grounded theory approaches. “Initiating rapport” was one of four categories developed from nurses and patients’ descriptions of the types of nursing care they experienced.

Initiating rapport

Rapport was established by informal, social communication that enabled the nurse and the patient to get to know each other as persons. One of the nurses interviewed described this interaction:

Just by introducing yourself, by chatting along as you’re doing things...with the patient. Asking them...questions about themselves...like ‘how are you feeling about being in hospital? How are you felling about the operation tomorrow?’ And then they’ll sort of give you a clue...and actually then tell you how they’re feeling about things...just general chit chat... (Nurse).

(Williams & Irurita, p. 38)

The sequence used by Williams and Irurita to describe key information about the category is an effective style of reporting qualitative findings. It consists of:

- A label for the category
- The authors’ description of the meaning of the category
- A quotation from the raw text to elaborate the meaning of the category and to show the type of text coded into the category.

This style is recommended for reporting the most important categories that comprise the main findings from an inductive analysis.

Williams and Irurita also included a figure including the four main categories derived from their grounded theory analysis that linked the categories into an overall model. These categories were: (1) *initiating rapport*, (2) *developing trust*, (3) *identifying patient needs* and (4) *delivering quality nursing care*. In the model these categories formed a sequence where each earlier step needed to be accomplished before the next step could develop (Williams & Irurita, 1998, p. 38).

Meta-ethnography of self-management of diabetes

In a meta-ethnography of qualitative studies that investigated on lay experiences of diabetes and diabetes care, Campbell and her colleagues developed a model that incorporated eight key concepts relating to self-care of diabetes. One of the major outcome categories they referred to as *strategic non-compliance*, which they described as

...the thoughtful and selective application of medical advice rather than blind adherence to it. The synthesis indicated that such an approach was associated

with being in control of diabetes, ‘coping’, achieving a balance between the quality of life and the illness, improved glucose levels and a feeling of well-being.” (Campbell et al, 2003 p. 680)

In a further elaboration they described *strategic non-compliance* as:

... a balanced life with diabetes was found to be strongly associated with an approach to the self-management of diabetes characterised as ‘strategic non-compliance’, involving the monitoring and observation of symptoms and an ability to manipulate dietary and medication regimens in order to live life as fully as possible, rather than limiting social and work activities in order to adhere rigidly to medical advice. (Campbell et al, 2003, p. 681)

These more extended descriptions convey the detailed meaning of the category “strategic non-compliance” which was one of the important themes emerging from the meta-ethnography.

One of the categories contributing to effective self-management of diabetes was labelled *acknowledgement diabetes is serious*. The text, which elaborated the meaning of this category, was:

- (a) ‘Serious but not for me’—those who felt that diabetes was a serious illness in general, but that they did not have the serious type.
- (b) ‘Serious but I can control it’—those who felt that diabetes could have serious effects but that these were controllable through radical behaviour changes.
- (c) ‘Serious for me’—those who felt pessimistic about prognosis, mainly due to perceived personal candidacy for complications, were described as ‘serious for me’. (Campbell et al, 2003, p. 680)

Summary points

An important point to note is that most inductive studies report between three and eight main categories in the findings (e.g., Campbell et al, 2003; Jain & Ogden, 1999). Inexperienced researchers often try to report 10 or more major categories. This indicates that they have probably not finished the process of combining the smaller categories into more encompassing categories or they have not made the crucial decisions about which categories are the most important.

The *general inductive approach* provides a convenient and efficient way of analysing qualitative data for many research purposes. The outcomes of analysis may be indistinguishable from those derived from a grounded theory approach. However, unlike grounded theory, there is no emphasis on learning new technical terms such as “open coding” and “axial coding.” For this reason, many researchers are likely to find using a general inductive approach more straightforward than some of the traditional approaches to qualitative data analysis.

References

- Backett, K. C., & Davison, C. (1995). Lifecourse and lifestyle: The social and cultural location of health behaviours. *Social Science & Medicine*, 40(5), 629-638.
- Bryman, A. & Burgess, R.G. (eds). (1994). *Analyzing qualitative data*. London: Routledge.
- Campbell, R., Pound, P., Pope, C., Britten, N., Pill, R., Morgan, M., et al. (2003). Evaluating meta-ethnography: a synthesis of qualitative research on lay experiences of diabetes and diabetes care. *Social Science & Medicine*, 56, 671-684.
- Creswell, J. W. (2002). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Upper Saddle River, NJ: Pearson Education.
- Dey, I. (1993). *Qualitative Data Analysis: A User-Friendly Guide for Social Scientists*. London: Routledge.
- Elliott, S. J., & Gillie, J. (1998). Moving experiences: a qualitative analysis of health and migration. *Health & Place*, 4(4), 327-339.
- Ezzy, D. (2002). *Qualitative analysis: Practice and innovation*. Crows Nest, NSW: Allen & Unwin.
- Erlandson, D. A., Harris, E. L., Skipper, B. L., & Allen, S. D. (1993). *Doing naturalistic enquiry: A guide to methods*. Newbury Park, CA: Sage.
- Jain, A. & Ogden, J. (1999). General practitioners' experiences of patients' complaints: qualitative study. *British Medical Journal*, 318, 1596-1599.
- Leiblich, A. (1998). *Narrative research: Reading, analysis and interpretation*. Thousand Oaks, CA: Sage.
- Marshall, M. N. (1999). Improving quality in general practice: qualitative case study of barriers faced by health authorities. *British Medical Journal*, 319, 164-167.
- McGregor, K. (2003). "Therapy - it's a two-way thing": Women survivors of child sexual abuse describe their therapy experiences. Unpublished doctoral thesis, Psychology Department, University of Auckland, New Zealand
- Miles, M. B. & Huberman, A. M. (1994). *Qualitative data analysis* (2nd ed.). London: Sage.
- Pope, C., Ziebland, S., & Mays, N. (2000). Qualitative research in health care: Analysing qualitative data. *British Medical Journal*, 320, 114-116.
- Potter, J., & Wetherall, M. (1994). Analyzing discourse. In A. Bryman & R. Burgess (Eds.), *Analyzing qualitative data* (pp. 47- 68). London: Routledge.
- Richards, T., & Richards, L. (1995). Using hierarchical categories in qualitative data analysis. In U. Kelle (Ed.), *Computer-aided qualitative data analysis: Theory, methods and practice* (pp. 80-95). London: Sage.

- Silverman, D. (2000). *Doing qualitative research: A practical handbook*. London: Sage.
- Stolee, P., Zaza, C., Pedlar, A., & Myers, A. M. (1999). Clinical experience with goal attainment scaling in geriatric care. *Journal of Aging and Health, 11*(1), 96-124.
- Strauss, A. & Corbin, J. (1990). *Basics of Qualitative Research*. Newbury Park: Sage.
- Van Manen, M. (1990). *Researching lived experience: Human science for an action sensitive pedagogy*. London, Ontario: Althouse Press.
- Williams, A. M., & Irurita, I. F. (1998). Therapeutically conducive relationships between nurses and patients: an important component of quality nursing care. *Australian Journal of Advanced Nursing, 16*(2), 36-44.