

Analysing Qualitative Data

Analysing qualitative data is an iterative process of alternating between emergent readings of the data and drawing in the use of existing models, explanation and theories. This is a reflexive process. The researcher visits and revisits the data, connects them to emerging insights, and progressively refines his/her focus and understandings (Srivastava & Hopwood, 2009, p. 77).

The iterative data analysis involves:

Preparing the data

- Format and label your data. Create contact sheets and lists of pseudonyms
- Gather the data to be analysed. Order them. (Re)label, print and reformat the data
- Decide which data (fieldnotes, interview transcripts, documents) to include in your analysis
- Organise your data systematically into folders – chronologically; by data type; or by source

Data immersion

When you are about three-quarters in your data collection, submerge yourself in the data:

- Read and re-read your data
- Talk to others about your data and emerging findings
- Jot down your reflections and hunches; consider a variety of interpretations, but hold off judgement
- Speak to your data. Ask, “What is happening here?”; “What strikes me?” (Creswell, 2007).
- Stay open to multiple meanings. Think, “Is there a story here?”.

Primary-cycle coding

- Also known as “**open coding**” and “**initial coding**” (Charmaz, 2006; Glaser & Strauss, 1967) and “first cycle coding” (Saldana, 2011). At this level of coding, the analyst opens up the meaning in the data. This coding process is also referred to a cycle to capture the circular reflexive process that marks qualitative data analysis. Thus, in primary-cycle coding, the initial coding activities of ‘breaking up’ the meanings in the data occur more than just a single “first” time. The data might be read and coded several times during the primary stage.
- Primary-cycle coding begins with an examination of the data and assigning words or phrases that capture their essence. If you use Microsoft Word, you can either write the codes into a right-hand margin or into a “new comment” bubble inside the Review tab.
- These first-level codes focus on “what” is present in the data, i.e. they are descriptive codes, showing the basic activities and processes in the data. Keep these codes as active and as specific as possible. Eg. Use gerunds (nouns ending with -ing. Eg. running; laughing) to help you do this.
- First-level codes which make use of the actual words or phrases within a segment of data are called *in vivo* codes.
- Throughout the coding process, qualitative researchers use the **constant comparative method** (Charmaz, 2006) to compare the data applicable to each code, and they modify code definitions to fit new data; or break them off and create a new code. This is a circular, iterative, and reflexive process.

- Note: The iterative approach does not require that the entire corpus of data be put through a fractured and detailed primary coding cycle. Run first-cycle coding on parts of the data that are typical or interesting or contrastive.

Focusing the analysis and creating a codebook

- Create a list of codes and a brief definition or representative example of each code.
- Streamline this start list (Miles & Huberman, 1994) of codes (30 - > 300) into a **codebook** – a data display that systematically lists key codes, definitions and examples that are going to be used in your analysis. Codebooks are crucial for team coding to ensure inter-coder reliability and that all analysts are working from the same 'legends' (definitions, examples, etc).
- The codebook can morph throughout the data analysis process. It serves as a chronological map registering how the codes merged and changed over time. Codebooks are also helpful for explaining the data analysis process to supervisors and reviewers.
- Keep the total number of codes below 25.
- Throughout the analysis, revisiting your research question and sensitizing concepts helps you ensure they are still relevant and interesting. After some primary-cycle coding, you should reconsider the best direction of the analysis, rework research questions and focus, and educate yourself on literature that frames new directions.

Secondary-cycle coding

The researcher critically examines the codes already identified in the primary cycle and begins to organize, synthesizes and categorize them into interpretive concepts. Codes here are '**focussed**' (Charmaz, 2006; Saldana, 2009). This involves identifying patterns, rules or cause-effect progressions:

- You group and link the second-level codes in a specific way – known as **axial coding** (Charmaz, 2006; Strauss & Corbin, 1998) or hierarchical codes (Tracy, 2013).
- At this stage, you decide which parts of your data should be put aside, and which should be pursued in order to flesh out an emerging code or explanation. This practice is called **theoretical sampling**. You know you have gathered enough data when new pieces add little, if any, new value to the emergent analysis – a state called **theoretical saturation** (Glaser & Strauss, 1967).

Synthesising and making meaning from codes

- Record your emerging analysis thoughts and ideas systematically. Your notes here will be useful for drafting your methods section in your paper later on.
- Write **analytic memos**. These are sites of conversation between the researcher and the data (Clarke, 2005) and a mind-dump of ideas (Saldana, 2009). Save the memos in a Word document or in a qualitative computer software system. Analytic memos are helpful for thinking through how codes relate to one another. Examine for antecedents and consequences of various codes. Reflecting on and making hypotheses about these linkages is crucial for understanding process, action, chronology, explanation and causation.
- You can also be your own devil's advocate by seeking out deviant data – **negative case analysis** – that do not appear to support the emerging hypothesis and revise your arguments so they better fit all the emerging data. Make a loose analysis outline that notes the primary research question and focus and the potential ways the emerging codes are attending to them. This outline can be used to lay out the plan of your paper.
- Make a **loose analysis outline** that notes the primary research question and focus and the potential ways the emerging codes are attending to them. This outline can be used to lay out the plan of your paper.

Table: Flow of Data Analysis Process

<p><i>Preparing the data</i></p>	<ul style="list-style-type: none"> • Format and label your data. Create contact sheets and lists of pseudonyms • Gather the data to be analysed. Order them. (Re)label, print and reformat the data • Decide which data (fieldnotes, interview transcripts, documents) to include in your analysis • Organise your data systematically into folders – chronologically; by data type; or by source
<p><i>Data immersion</i></p>	<ul style="list-style-type: none"> • When you are about three-quarters in your data collection, submerge yourself in the data: <ul style="list-style-type: none"> • Read and re-read your data • Talk to others about your data and emerging findings • Jot down your reflections and hunches; consider a variety of interpretations, but hold off judgement • Speak to your data. Ask, “What is happening here?”; “What strikes me?” (Creswell, 2007). • Stay open to multiple meanings. Think, “Is there a story here?”.
<p><i>Primary cycle coding</i></p>	<ul style="list-style-type: none"> • Also known as “open coding” and “initial coding” (Charmaz, 2006; Glaser & Strauss, 1967) and “first cycle coding” (Saldana, 2011). At this level of coding, the analyst opens up the meaning in the data. This coding process is also referred to a cycle to capture the circular reflexive process that marks qualitative data analysis. Thus, in primary-cycle coding, the initial coding activities of ‘breaking up’ the meanings in the data occur more than just a single “first” time. The data might be read and coded several times during the primary stage. • Primary-cycle coding begins with an examination of the data and assigning words or phrases that capture their essence. If you use Microsoft Word, you can either write the codes into a right-hand margin or into a “new comment” bubble inside the Review tab. • These first-level codes focus on “what” is present in the data, i.e. they are descriptive codes, showing the basic activities and processes in the data. Keep these codes as active and as specific as possible. Eg. Use gerunds (nouns ending with -ing. Eg. running; laughing) to help you do this. • First-level codes which make use of the actual words or phrases within a segment of data are called in vivo codes. • Throughout the coding process, qualitative researchers use the constant comparative method (Charmaz, 2006) to compare the data applicable to each code, and they modify code definitions to fit new data; or break them off and create a new code. This is a circular, iterative, and reflexive process. • Note: The iterative approach does not require that the entire corpus of data be put through a fractured and detailed primary coding cycle. Run first-cycle coding on parts of the data that are typical or interesting or contrastive.
<p><i>Focusing the analysis</i></p>	<ul style="list-style-type: none"> • Create a list of codes and a brief definition or representative example of each code. • Streamline this ‘start list’ (Miles & Huberman, 1994) of codes (30 - > 300) into a codebook – a data display that systematically lists key codes, definitions and examples that are going to be used in your analysis. Codebooks are crucial for team coding to ensure inter-coder reliability and that all analysts are working from the same ‘legends’ (definitions, examples, etc). • The codebook can morph throughout the data analysis process. It serves as a chronological map registering how the codes merged and changed over time. Codebooks are also helpful for explaining the data analysis process to supervisors and reviewers. • Keep the total number of codes below 25. • Throughout the analysis, revisiting your research question and sensitizing concepts helps you ensure they are still relevant and interesting. After some primary-cycle coding, you should reconsider the best direction of the analysis, rework research questions and focus, and educate yourself on literature that frames new directions.



<p><i>Secondary cycle coding</i></p>	<ul style="list-style-type: none"> The researcher critically examines the codes already identified in the primary cycle and begins to organize, synthesizes and categorize them into interpretive concepts. Codes here are 'focussed' (Charmaz, 2006; Saldana, 2009). This involves identifying patterns, rules or cause-effect progressions: <ul style="list-style-type: none"> You group and link the second-level codes in a specific way – known as axial coding (Charmaz, 2006; Strauss & Corbin, 1998) or hierarchical codes (Tracy, 2013). At this stage, you decide which parts of your data should be put aside, and which should be pursued in order to flesh out an emerging code or explanation. This practice is called theoretical sampling. You know you have gathered enough data when new pieces add little, if any, new value to the emergent analysis – a state called theoretical saturation (Glaser & Strauss, 1967).
<p><i>Synthesising</i></p>	<ul style="list-style-type: none"> Record your emerging analysis thoughts and ideas systematically. Your notes here will be useful for drafting your methods section in your paper later on. Write analytic memos. These are sites of conversation between the researcher and the data (Clarke, 2005) and a mind-dump of ideas (Saldana, 2009). Save the memos in a Word document or in a qualitative computer software system. Analytic memos are helpful for thinking through how codes relate to one another. Examining for antecedents and consequences of various codes. Reflecting on and making hypotheses about these linkages is crucial for understanding process, action, chronology, explanation and causation. You can also be your own devil's advocate by seeking out deviant data – negative case analysis – that do not appear to support the emerging hypothesis and revise your arguments so they better fit all the emerging data. Make a loose analysis outline that notes the primary research question and focus and the potential ways the emerging codes are attending to them. This outline can be used to lay out the plan of your paper.



An example of first-cycle (initial or open) coding using gerunds

In-vivo	Initial Codes
<p>Interviewer: Welcome to our focus group on communicating with patients. My first question is what are the challenges and difficulties that you encounter in your daily ward work?</p> <p>Junior Doctor: One challenge is speaking to families about withdrawing treatment for what we think are medically futile patients because some families feel that we are not giving the patient a chance and think that we are killing the patient and it is very hard to bring across the message that we are trying to improve his quality of life by not doing certain treatments. It's trying to convince them that whatever you are doing for the patient is for their best interest and not like we are trying to kill the patient or not giving the patient a chance.</p>	<p>Speaking to families about withdrawing treatment</p> <p>Withdrawing treatment for medically futile patients</p> <p>Patient's families feeling that doctors are not giving the patient a chance</p> <p>Patient's family thinking that doctors are harming / killing the patient</p> <p>Experiencing difficulty in explaining to patient's family</p> <p>Withholding treatment to improve quality of life</p> <p>Experiencing difficulty in convincing patient's family that withholding treatment is best for the patient</p> <p>Trying to convince patient's family that the doctors are working for the patient's best interest</p> <p>Trying to convince patient's family that the doctors are not harming the patient</p> <p>Trying to convince patient's family that one is not denying patient a chance to live</p>

What may be some second-level (focussed or axial) codes?

Glossary of key terms for analysing themes in qualitative data

analytic memos sites of conversation between the investigator and their data (Clarke, 2005) and a mind-dump of the ongoing investigation (Saldana, 2009).

codebook a type of data display or legend that lists key codes, definitions and examples that are going to be used in the analysis.

codes words or short phrases that capture a “summative, salient, essence-capturing and/or evocative attribute for ... language-based or visual data” (Saldana, 2009, p.3).

coding active process of labelling and systematizing data as belonging to or representing some aspects of a phenomenon. This phenomenon may be a concept, belief, action, theme, cultural practice or relationship. You can code using paper and coloured pens and highlighters, on an Excel spreadsheet or a Word document or a computer-aided qualitative data analysis software.

constant comparative method a method of analysis used to compare data applicable to each code and to modify code definitions so as to fit new data. Or else break off and create a new code. See Charmaz (2006).

data immersion phase a phase of data analysis during which researchers read and re-read their data, talk with others about them, and marinate in the emerging findings.

first-level code a type of code that is descriptive, shows the data’s basic content and processes, requires little interpretation, and focuses on “what” is present in the data.

in vivo codes ‘live’ quotes employing language and terms used by the participants themselves. Marked by “ ” quotation marks.

loose analysis outline an outline that notes the primary research questions and potential ways in which the emergent codes are attending to these questions.

negative case analysis seeking out deviant data that do not appear to support the emerging hypothesis, and revising arguments so that they fit all the emerging data better.

primary-cycle coding initial coding activities, which begin by examining the data and assigning words or phrases that capture their essence.

secondary-cycle coding critical examination of the codes already identified in primary cycles; at this stage the researcher begins to organize, synthesize and categorize these codes into interpretive and sometimes disciplinary concepts.

second-level codes codes that serve to explain, theorize, and synthesize the data; they include interpretation and help the researcher identify patterns, rules, or cause-effect progressions.

theoretical sampling strategic approach to answer the research question and to gain knowledge. Informs the research design.

theoretical saturation a state in which new data add little, if any, new value to the emergent analysis.

Practice Points

- Given the data you have collected, read and coded thus far, what are some interesting themes or issues?
 - a. Do these themes meaningfully intersect with the literatures and theories that you are already acquainted with? How so?
 - b. In what ways do these themes intersect with literatures and theories that pair well with qualitative methods?
 - c. In what ways do these themes intersect with literatures or theories that you are unfamiliar with, but am drawn to and willing / have time to learn more about?

- When you write memos,
 - a. Define the code as carefully as possible
 - b. Explicate its properties
 - c. Provide examples of raw data that illustrate the code
 - d. Specify conditions under which it arises, is maintained, and changes
 - e. Describe its consequences
 - f. Show how it relates to other codes
 - g. Develop hypotheses about the code

On the basis of the memos, develop and discuss one or more primary claims that may frame your analysis. What data must you still collect in order to examine the strength and tenability of these claims?

Resources

The information above is based on chapter nine of Tracy, S. (2013). *Qualitative research methods: Collecting evidence, crafting analysis, communicating impact*. Wiley-Blackwell.

The other readings mentioned here are:

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. Sage.

Clarke, A. E. (2005). *Situational analysis: Grounded theory after the postmodern turn*. Sage.

Glaser, B.G., & Strauss, A.L. (1967). *The discovery of grounded theory*. Aldine de Gruyter.

Miles, M.B., & Huberman, A.M. (1994). *Qualitative data analysis*. Sage.

Saldana, J. (2009). *The coding manual for qualitative researchers*. Sage.

Srivastava, P., & Hopwood, N. (2009). A practical literature framework for qualitative data analysis. *International Journal of Qualitative Methods*, 8, 76-84.

Strauss, A.L., & Corbin, J. (1998). *Basics of qualitative researcher: Techniques and procedures for developing grounded theory* (2nd ed.). Sage.