Systematic Data Analysis in Qualitative Health Research: Building Credible and Clear Findings

Stephen Kodish, Joel Gittelsohn
Johns Hopkins Bloomberg School of Public Health, Department of International Health, Social & Behavioral Interventions Program & Center for Human Nutrition, Baltimore, MD, USA

Overview
Textual data sets can be intimidating to public health researchers and practitioners who are unfamiliar with qualitative research. The amount of textual data collected from in-depth interviews (IDI), focus group discussions (FGD), and direct observations – three common methods in qualitative research – can be extensive and can prove challenging to systematically analyze.

This article outlines one primary approach to qualitative data analysis (QDA) in health research and discusses the analysis process and interpretation, leading to the development of a credible product. It also briefly describes how computer software can assist in both the analysis and display of findings through data graphs, tables, and conceptual models.

A dynamic aspect of qualitative inquiry is the variety of approaches that one can take. The disparate approaches include, but are not limited to, phenomenology, ethnography, or grounded theory, each of which might utilize a different analytic approach to data. Although there is no one-size-fits-all approach to data analysis in qualitative health research, commonalities across methodological approaches do exist and can be represented by an illustrative schemata (Figure 1) developed by Creswell.

Analysis starts at the bottom of the figure (i.e., during data collection) and proceeds upward through various stages until a written account is developed that presents the findings. The spiral image highlights the non-linear, iterative nature of QDA and offers both procedures and examples throughout each stage of the process, from initial data management to representation of findings. The remainder of this paper will highlight the three procedural stages at the top of Creswell’s spiral: “Reading, Memoing”, “Describing, Classifying, Interpreting”, and “Representing, Visualizing”.

Reading, memoing
An important analytic strategy in QDA is memo writing, or “memoing”, which assists a researcher in making a conceptual bridge from raw textual data to abstractions used to explain the phenomena of interest. It is the process of writing down thoughts and questions in relation to the text in which the researcher is immersed. Writing memos is often an intermediate step between data collection and coding and, as Charmaz explains, the process of memoing helps to, “catch your thoughts, capture the comparisons and connections you make, and crystallize questions and directions you want to pursue.”

Describing, classifying, interpreting
After textual data have been collected, read, and reviewed, a researcher may begin coding the data in order to reduce them into meaningful segments for interpretation. Any kind of textual data can be coded, including memos, field notes, or direct observation notes. This article focuses on in-depth interview and focus group discussion data due to their popularity as methods in the field. Coding is a process of identifying themes – that is, analytic categories – in text and is one of the key elements in QDA. Codes, identifiers of themes in the coding process, are the building blocks for theory or model building and the foundation on which project findings most often rest.

One might develop 100 codes for a data set or perhaps just 10. For ease of interpretation and clarity, however, Creswell recommends utilizing no more than 25-30 categories of information regardless of the size of the database. Coding can be inductive or deductive.

Inductive coding
An inductive coding approach is commonly utilized in an early, exploratory stage of a research project, when the researcher has
not formulated hypotheses, and is based on few (if any) pre-conceived notions of the final results. Plans for additional data collection are frequently the outcome of early coding with this exploratory approach.

Consider, for instance, a qualitative acceptability study of a specialized food commodity, such as a lipid-based nutrient supplement (LNS), perhaps Nutributter™ (Nutriset SAS, Malaunay, France). LNS containing energy, protein, essential fatty acids, and micronutrients have been developed to overcome nutrient shortfalls in existing diets of young children 6–24 months.⁹ Although Nutributter™ has been accepted by target populations in some settings,¹¹ a researcher examining acceptability in a new setting using a qualitative approach might analyze initial qualitative data using open coding; that is, he or she explores the textual data line-by-line for conceptualization of “emergent” themes related to beneficiary perceptions of the unfamiliar commodity.³⁻⁵ Themes from the data might emerge that are unexpected to the researcher, for example, unique cultural characteristics that directly relate to acceptability. In such a scenario, analysis using inductive coding would be concurrent with data collection to shape future stages of the research project. New questions could be asked or additional methods added based on those emergent themes.

**Deductive coding**

Deductive coding is oriented towards confirming or testing the investigator’s preconceived terms and relationships. It is often based on a researcher’s a priori⁴ (a Latin term that refers to prior knowledge about a population) hypotheses and might utilize “prefigured” codes derived from a theoretical model or existing literature on the topic of interest.¹²,¹³ Using pre-determined codes is popular in the health sciences – a field that utilizes many models to explain health-seeking behavior.

As an example, consider the Theory of Planned Behavior (TPB) (Figure 2), which seeks to explain why people perform certain actions¹⁴,¹⁵ – for example adhering to daily consumption of Nutributter™. A researcher deductively analyzing a textual data set would apply codes based on the TPB in relation to the major constructs of the theory: perceived behavioral control (PBC),

---

**FIGURE 1: Data analysis spiral**

![Data analysis spiral diagram](image1)

**FIGURE 2: Visual representation of the theory of planned behavior**

![Visual representation of the theory of planned behavior](image2)
attitude (ATT), subjective norm (SN), and intention (INT). While examining texts, he or she would specifically look for themes in relation to those constructs and, perhaps, ignore other topics unrelated to the theory. Creswell suggests, however, that researchers who employ this approach to analysis be open to additional codes emerging during the analytic process, because using this type of coding scheme may limit the analysis to the “prefigured” codes rather than open up the codes to reflect the views of participants from an *emic* perspective and, consequently, may limit findings.

**Using a codebook**

In both deductive and inductive coding, researchers usually develop a codebook to assist with the process. The standardized structure of a codebook provides a stable frame for the analysis of textual data and can help to establish more stability and guidance when coding. Put simply, a codebook is a reference tool that tells a researcher when to apply what code to a chunk of text in a transcript. Both the codes themselves and their respective definitional parameters should be included in a codebook.

**Representing, visualizing**

Following memoing and coding, researchers present what was found during analysis, often in the form of a table, matrix, or chart. A visual representation of findings can be helpful for summarizing and highlighting key findings. For example, a simple 2 x 2 table that compares individuals by gender or ethnic group in terms of one of the themes or categories in the study might be useful and informative. In public health research, conceptual

**TABLE 1:** Excerpt of a codebook developed for inductive QDA of data collected at Kakuma Refugee Camp, Kenya.

<table>
<thead>
<tr>
<th>Mnemonic or numeric “Brief” Code</th>
<th>Full Description of Code</th>
<th>When to use/not to use the code</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0 Life in Kakuma</td>
<td>Refugee experiences residing in KRC</td>
<td>Use this family of codes when the CL or MNP beneficiary discusses his or her life as a refugee at KRC.</td>
</tr>
<tr>
<td>2.1 Hardship</td>
<td>Hardships faced while living in Kakuma, related to security, violence, tribalism, etc</td>
<td>Use this code for the array of hardships refugees discuss at KRC unrelated to illness experiences. Illness is mentioned a lot but use 2.2.</td>
</tr>
<tr>
<td>2.2 Illness</td>
<td>Illness experiences of the individual or of his or her family and/or community</td>
<td>Use this umbrella code for any health-related experience related to life in KRC. It can be related to anemia or another illness. Codes 2.2.1 and 2.2.2 will be used to distinguish between the types of illness discussed.</td>
</tr>
<tr>
<td>2.2.1 Ill. Anemia</td>
<td>Experiences with anemia or malnutrition, specifically</td>
<td>Use this code for health-related experiences, in particular those related to anemia and/or malnutrition. Also, “lack of blood” should be included in this code as it’s referring to anemia.</td>
</tr>
</tbody>
</table>

**TABLE 2:** Advantages and disadvantages of using computer software.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Provides an organized storage file system, especially large data sets</td>
<td>1. Requires the researcher to learn how to use the software program, which can be very time consuming</td>
</tr>
<tr>
<td>2. Helps a researcher locate textual material quickly</td>
<td>2. Better programs may be cost prohibitive</td>
</tr>
<tr>
<td>3. Creates visually informative schemata to illustrate findings</td>
<td>3. Distances the researcher from the data</td>
</tr>
<tr>
<td>4. Provides time-saving functions (e.g., quickly can determine frequencies of codes)</td>
<td>4. Makes analysis as a team challenging due to logistics behind sharing files</td>
</tr>
<tr>
<td>5. Allows for coding of not only text, but also images and video files</td>
<td>5. Nascent compared to quantitative software programs can be frustrating to work with</td>
</tr>
</tbody>
</table>

Content adapted from and the authors’ experiences

3 Put simply, a codebook is a reference tool that tells a researcher when to apply what code to a chunk of text in a transcript. Both the codes themselves and their respective definitional parameters should be included in a codebook.

8 In public health research, conceptual
models are commonly used to illustrate relationships between themes, with the purpose of showing how different factors relate to a health-seeking behavior or outcome. Figure 3 is offered as an example.16 When choosing the most appropriate display of findings, one should consider not only what visual representation most clearly and completely answers the research question(s), but also the audience for whom the graphics are intended.

Using computer software

Computer software programs are available to help with the analysis and presentation of textual data. Such programs help a researcher code and retrieve text, create data matrices, and build models of how the themes in a data set are associated with each other.9 As the process used for textual analysis is the same for hand coding or using a computer, this type of software may be most useful while working with large data sets (e.g., more than 500 pages of text3) and an unnecessary burden while working with those smaller. (Table 2) Two popular commercial programs available include NVivo (Nvivo (Version 9.0). [Computer Software]. Victoria, AU: QSR International Pty Ltd) and Atlas.ti (Atlas.ti (Version 6.1) [Computer Software]. Berlin: Scientific Software Development), both of which can only be used with Windows-based operating systems.

Credibility of qualitative research findings

Computer software can help with coding, but it cannot help ensure high-quality data or findings. Strategies exist to enhance the quality in qualitative research, some during data collection and others during analysis. Creswell points to eight major strategies that can be utilized to help ensure the “trustworthiness” of a study and recommends that at least two be used in any given qualitative study. (Table 3) Member checking, peer debriefing, and investigator triangulation, described in Table 3, are particularly useful tools for enhancing the credibility of qualitative findings.

Conclusions

Creswell’s data analysis spiral offers a good representation of the dynamic process that QDA should undergo. It highlights an iterative and systematic approach to data analysis that can help to ensure credible findings.17 However, just as is the case while analyzing a quantitative data set, one’s findings are only as good...
as the data that have been collected. Methodological rigor during data collection can help make analysis easier and findings more credible.

**Correspondence:** Stephen Kodish, MS, Social & Behavioral Interventions Program, Department of International Health, The Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe St., Baltimore, MD 21205-2130, USA  
**Email:** skodish@jhsph.edu

**References**


**Table 3: Qualitative data validation procedures.**

<table>
<thead>
<tr>
<th>Validation Procedure</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Prolonged engagement in the field</td>
<td>Building trust with participants, learning the culture, and checking for misinformation that stems from distortions introduced by the research team</td>
</tr>
<tr>
<td>2. Triangulation</td>
<td>Making use of multiple and different sources, methods, investigators, and theories for data corroboration</td>
</tr>
<tr>
<td>3. Peer review</td>
<td>An external check of the research process</td>
</tr>
<tr>
<td>4. Searching for negative cases</td>
<td>Refinement of a working hypothesis by an active search for disconfirming evidence</td>
</tr>
<tr>
<td>5. Clarifying researcher bias</td>
<td>Critically reflecting on what the researcher, him or herself, brings to the research project (eg, past experiences, prejudices, etc)</td>
</tr>
<tr>
<td>6. Member checking</td>
<td>Soliciting participants’ views of the credibility of the findings and interpretations during analysis</td>
</tr>
<tr>
<td>7. Providing a thick description</td>
<td>Enables the reader to determine which characteristics, if any, of a program can be transferred to other settings through a detailed description of participants and setting</td>
</tr>
<tr>
<td>8. External audits</td>
<td>When an external consultant examines both the process and product for accuracy</td>
</tr>
</tbody>
</table>

Source: Adapted from 3