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.[1]

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,[2] each of which might utilize a different analytic approach to data.[3] 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.[3]

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.[4] 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[5] 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."

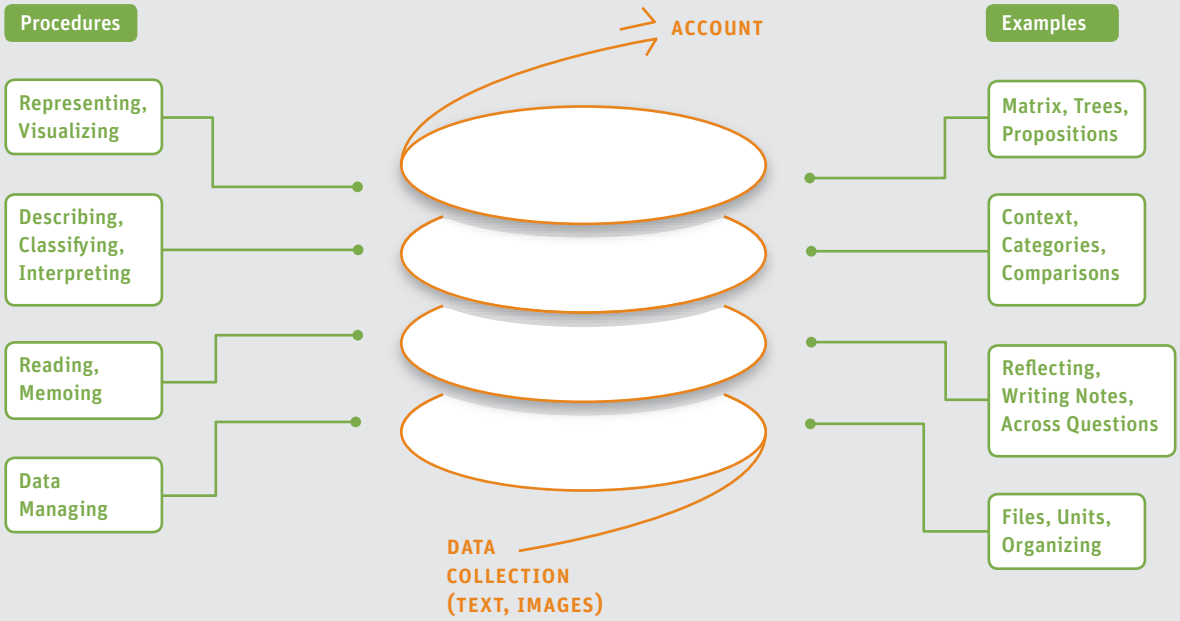
Describing, classifying, interpreting

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.[6,7] 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.[8] 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.[3] Coding can be inductive or deductive.[9]

Inductive coding

An inductive coding approach is commonly utilized in an early, exploratory stage of a research project, when the researcher has

FIGURE 1: Data analysis spiral

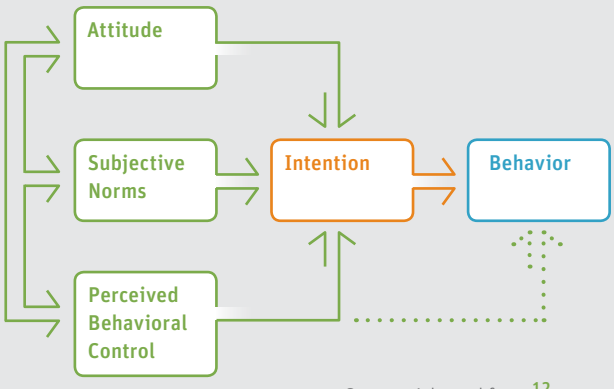


Source: Creswell [3]

not formulated hypotheses, and is based on few (if any) preconceived 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.[10] 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.

FIGURE 2: Visual representation of the theory of planned behavior



Source: Adapted from [12]

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 prior*[4] (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 actions[14,15] – 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),

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

Mnemonic or numeric "Brief" Code	Full Description of Code	When to use/not to use the code
2.0 Life in Kakuma	Refugee experiences residing in KRC	Use this family of codes when the CL or MNP beneficiary discusses his or her life as a refugee at KRC.
2.1 Hardship	Hardships faced while living in Kakuma, related to security, violence, tribalism, etc	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.
2.2 Illness	Illness experiences of the individual or of his or her family and/or community	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.
2.2.1 Ill. Anemia	Experiences with anemia or malnutrition, specifically	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.

Source: Adapted from one of the authors' projects

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[3] and, consequently, may limit findings.

TABLE 2: Advantages and disadvantages of using computer software.

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

Content adapted from [3,8] and the authors' experiences

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.[8] 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. **(Table 1)** In general, during inductive analysis a researcher develops the codebook as part of the coding process, whereas during deductive analysis the researcher develops the codebook before the textual analysis.

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 2x2 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.[6] 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 (eg, more than 500 pages of text[3]) 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.

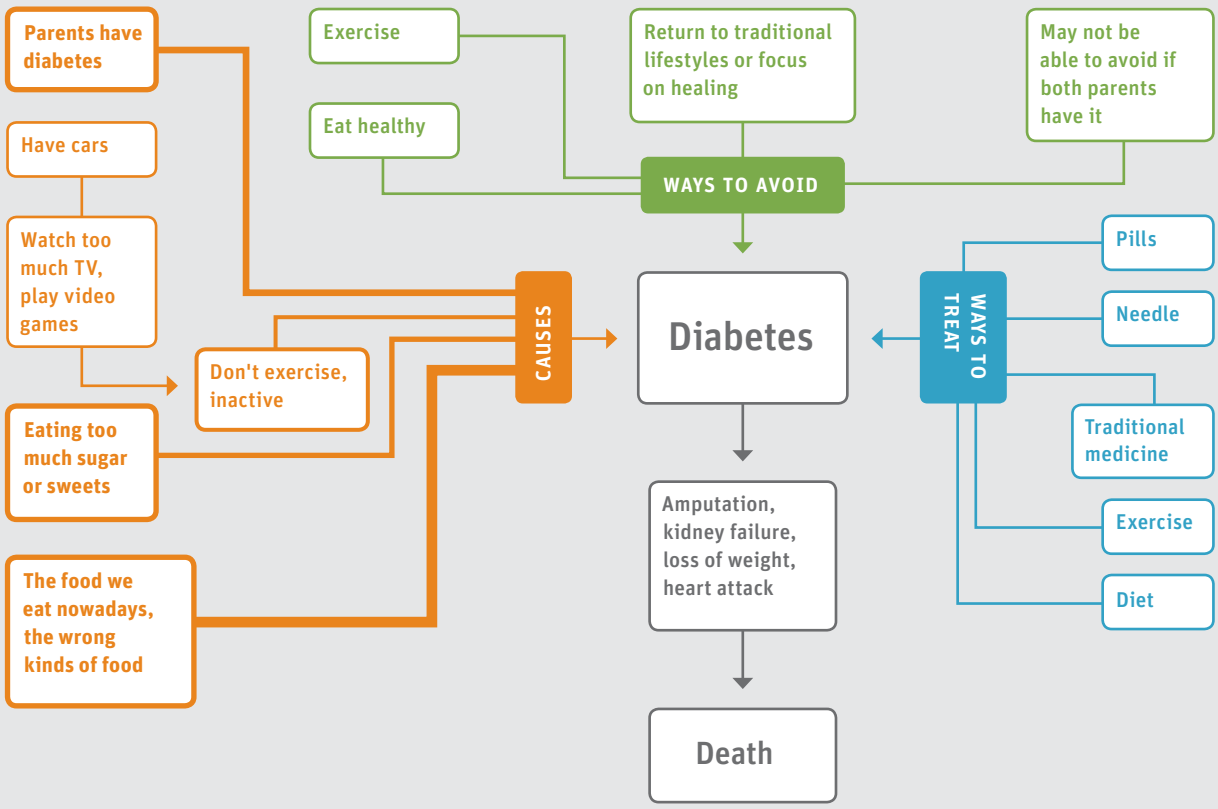
Credibility of qualitative research findings

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)**[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

FIGURE 3: Conceptual model illustrating findings from a diabetes prevention study



Source: Adapted from [13]

TABLE 3: Qualitative data validation procedures.

Validation Procedure	Explanation
1. Prolonged engagement in the field	Building trust with participants, learning the culture, and checking for misinformation that stems from distortions introduced by the research team
2. Triangulation	Making use of multiple and different sources, methods, investigators, and theories for data corroboration
3. Peer review	An external check of the research process
4. Searching for negative cases	Refinement of a working hypothesis by an active search for disconfirming evidence
5. Clarifying researcher bias	Critically reflecting on what the researcher, him or herself, brings to the research project (eg, past experiences, prejudices, etc)
6. Member checking	Soliciting participants' views of the credibility of the findings and interpretations during analysis
7. Providing a thick description	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
8. External audits	When an external consultant examines both the process and product for accuracy

Source: Adapted from[3]

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

01. Kodish S, Gittelsohn J. Report from the field: Exploring barriers to micronutrient powder uptake at Kakuma refugee camp. ***Sight and Life*** 2010;3:61-63.
02. Crotty M. The foundations of social science: meaning and perspective in the research process. Thousand Oaks: Sage Publications, 1998.
03. Creswell JW. Qualitative inquiry and research design: choosing among five approaches. 2nd ed. Thousand Oaks: Sage Publications, 2007.
04. Birks M, Chapman Y, Francis K. Memoing in qualitative research: probing data and processes. J Res Nurs 2008;13:68-75.
05. Charmaz K. Constructing grounded theory: a practical guide through qualitative analysis. Thousand Oaks: Sage Publications, 2006.
06. Miles MB, Huberman AM. Qualitative data analysis. 2nd ed. Thousand Oaks: Sage Publications, 1994.
07. Strauss A, Corbin J. Basics of qualitative research: grounded theory procedures and techniques. Newbury Park: Sage Publications, 1990.
08. MacQueen KM, McLellan E, Kay K et al. Codebook development for team-based qualitative analysis. Cult Anthro Meth 1999;10:31-36.
09. Bernard, HR. Research methods in anthropology: qualitative and quantitative approaches. 4th ed. Oxford: AltaMira Press, 2006.
10. Hess SY, Bado L, Aaron GJ et al. Acceptability of zinc-fortified, lipid-based nutrient supplements (LNS) prepared for the young children in Burkina Faso. Mat Ch Nutr 2010;1-11.
11. Adu-Afarquah S, Lartey A, Brown KH et al. Home fortification of complementary foods with micronutrient supplements is well accepted and has positive effects on infant iron status in Ghana. Am J Cl Nutr 2008;87:929-938.
12. Crabtree BF, Miller WL. Doing qualitative research. Newbury Park: Sage Publications, 1992.
13. Marshall C, Rossman GB. Designing qualitative research. 4th ed. Thousand Oaks: Sage Publications, 2006.
14. Ajzen I. Action to control: from cognition to behavior. In: Kuhi J, Beckmann J. eds. From intentions to actions: A theory of planned behavior. Heidelberg: Springer, 1985.
15. Ajzen I, Madden TJ. Prediction of goal-directed behavior: attitudes, intentions, and perceived behavioral control. J Exp Soc Psych 1986;22:453-474.
16. Ho LS, Gittelsohn J, Harris SB et al. Development of an integrated diabetes development program with First Nations in Canada. Hlth Prom Intl 2006;21:88-97.
17. Lincoln Y, Guba E. Naturalistic inquiry. New York: Sage, 1985.