Qualitative Methods: Coding & Data Analysis

Michele Andrasik, Ph.D.
Sarah Frey, MSW
Meheret Endeshaw, MPH
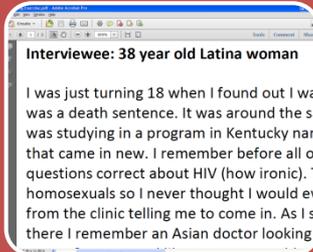
CFAR SPRC Qualitative Methods Workshop Series

Outline for Today

1. Data Management
2. Coding
3. Data Analysis

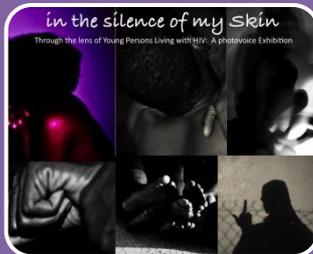# What does Qualitative Data look like?

Text

- Transcriptions of interviews & focus groups
- Notes & memos

Audio

- Audio recording

Visual

- Video
- Photograph

Data Management: Security

- How and where will you **store** your data?
  - Paper copies vs. digital copies
  - Teams: how will you share data while retaining security?
  - How and where will you back up your data?
- Removal of names and identifying characteristics on transcriptions (Anonymization) as early as possible. Destroy non-anonymized data (i.e. audio).

Data Management: Software

Software for qualitative analysis:

- **Atlas.ti ($99 student, $670 full)**
- **Nvivo ($215 student, $670 full)**
- **CDC's EZ-text ($0)**
- wikipedia **computer assisted qualitative data analysis software**: list of proprietary and open source software

Do you need software to do qualitative research?

Atlas.ti

Available at: www.atlasti.com

Nvivo

Available at: www.qsrinternational.com/

Definitions

Codes

short hand notation for themes that you see in the data

Coding

the act of linking themes/codes with passages of qualitative data

Codebooks

lists of codes and definitions of codes

Qualitative Researcher Qualities

Organization

Perseverance

Accept Ambiguity

Flexibility

Creativity

Ethical

Vocabulary

Tips for Beginners

Code on hard copy printouts first –

Not via a computer monitor

Once codes are fairly well set then **transfer codes to electronic file**

Begin working with qualitative software *AFTER* you have understanding of fundamentals of qualitative analysis

Lumping

- "Holistic coding" – broad brush strokes

- Very expedient

- Future detailed sub coding still possible

Pro	Con
Gets to essence of categorization of a phenomenon	May lead to superficial analysis if you do not employ conceptual words and phrases

Splitting

- □ "In vivo coding" – splitting data into smaller "codable" moments

- □ Generates more nuanced analysis from the beginning

Pro	Con
Encourages careful scrutiny of data	May overwhelm data analysis when it comes time to categorize codes

Coding

First Cycle Coding

- **Attribute Coding**
- **Descriptive Coding**
- **Structural Coding**
- **In Vivo Coding**
- **Analytic Coding**
- **Process Coding**
- **Topic Coding**

Second Cycle Coding

- **Pattern Coding**
- **Focused Coding**
- **Axial Coding**
- **Theoretical Coding**
- **Longitudinal Coding**

2ⁿᵈ Cycle Coding (Lumpers vs. Splitters)

Code 1

Code 2

Code A

Code 3
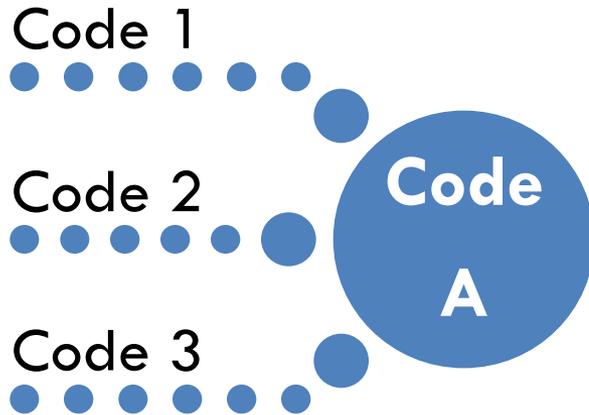
Collapse original number of 1ˢᵗ cycle codes into smaller numbers

- Reanalyze data
- Find larger segments of text are better suited to just one key code rather than several smaller ones

Code A

Code 1

Code B

Code C
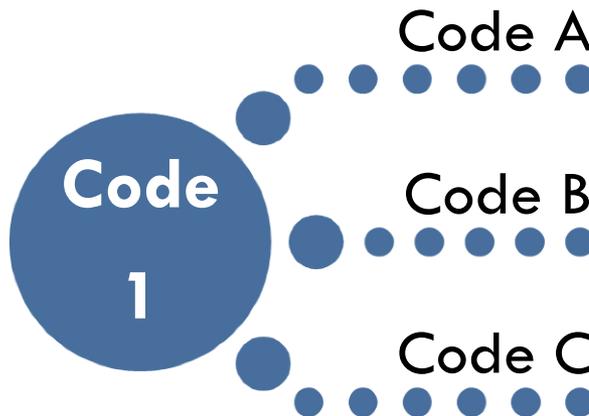
Expand number of 1ˢᵗ cycle codes into larger number of codes

- Reanalyze data
- Find segments of text are better suited in smaller codes rather than just one.

Example

"We think that sometimes parents, we don't talk about sex to our daughters. Therefore when they start to have sexual relations, they don't have as much knowledge on how to use a condom and that puts them at higher risk. They say that girls have a higher risk because there is less information about sex."

Types of Coding

Descriptive/Structural Coding:

- Describe characteristics of the data itself
- Answers who, what where, and how the data were collected.

Topic/Thematic Coding:

- Most common kind of coding
- Coding to describe topic; any passage will include several topics
- Creating a category or recognizing one from earlier

Analytic coding:

- Going beyond gathering by topic to analysis
- Pursue comparisons
- Ask yourself " What is this all about?"

Descriptive Codes

"We think that sometimes parents, we don't talk about sex to our daughters. Therefore when they start to have sexual relations, they don't have as much knowledge on how to use a condom and that puts them at higher risk. They say that girls have a higher risk because there is less information about sex."

- Father, 40 years old
- Interviewed at UW Medical Clinic
- Question: Why do you think people don't protect themselves from HIV?

Topic Codes

"*We think that sometimes* **parents, we don't talk about sex to our daughters**. *Therefore* **when they start to have sexual relations, they don't have as much knowledge on how to use a condom and that puts them at higher risk**. *They say that* **girls have a higher risk because there is less information about sex.**"

Topic codes:

- **Parent–child communication**
- **Lack of knowledge**
- **Vulnerability of girls**

Analytic Codes

We think that sometimes parents, we don't talk about sex to our daughters. Therefore when they start to have sexual relations, they don't have as much knowledge on how to use a condom and that puts them at higher risk. They say that girls have a higher risk because there is less information about sex."

- Make new categories based on what you see across interviews.
- Revisit categories to see if they still fit or change definition as you add more data.
- **Ex:** Do parents who feel comfortable talking about HIV with their children have <u>different</u> thoughts than parents who feel uncomfortable talking about HIV with children?

CODING EXERCISE

Discussion

□ What codes did you develop for your research question?

□ What would your codebook look like?

□ What was the experience like to code with other people and to reach consensus about codes?

How many codes are enough?

- Lichtman (2006) – Generate 80-100 codes that will be organized into 15-20 categories which eventually synthesize into five to seven (5-7) major concepts
- Cresswell (2007) – Begin with a short-list of five to six provisional codes to begin the process of "lean coding". This expands to no more than 25-30 categories that then combine into five to six (5-6) major themes
- Wolcott (1994) – 3 of anything major is a good quantity for reporting qualitative work
- Final number of themes/concepts should be kept to a minimum to keep analysis coherent

No Magic or standardized number to achieve.

The Codebook is KEY

☐ Create a codebook from the beginning—record of your emergent codes

☐ Review periodically – maintaining list provides opportunity to organize and reorganize codes into major categories and subcategories

Codebook Contents

- <u>codes</u>

- <u>definition</u> or content description

- guidelines for <u>when to use</u> the code

- guidelines for <u>when not to use</u> the code

- <u>examples</u>

Code	MARGIN
Definition	Marginalized members of the community. Groups that are perceived negatively as social and/or physical outsiders from the main community structure.
When to Use	This code applies to all references to groups of people who have been marginalized from the rest of the community.
When Not to Use	Don't use this code to refer to groups that are institutionalized for health or criminal justice reasons (see INSTIT).
Example	"…."

Team coding

☐ Pros/Cons to coding as a team

☐ Should I calculate inter-coder reliability?

Using Notes/Memos

- ☐ Making memos or notes of ideas/thoughts/ observations as you are coding (or immediately after an interview/FG)

- ☐ Can use to go back and make comparisons, deeper analysis

- ☐ Can treat as additional piece of data and code as you are coding the transcribed interview

References

- Cresswell, J.W. (2007). *Qualitative inquiry and research design: Choosing among five approaches* (2nd ed.). Thousand Oaks, CA: Sage

- Lichtman, M. (2006) *Qualitative research in education: A user's guide*. Thousand Oaks, CA: Sage

- Saldana, J. (2009). *The coding manual for qualitative researchers*. Thousand Oaks, CA: Sage

- Wolcott, H.F. (1994). *Transforming qualitative data: Description, analysis, and interpretation*. Thousand Oaks, CA: Sage.

Thank You!