# Qualitative Data Analysis & Coding

2018 USAID/JORDAN MEL CONFERENCE

INTERNATIONAL REPUBLICAN INSTITUTE

# BEFORE WE BEGIN…

…tell me about your experiences doing research!

- What did you especially like?

- What didn't you like?

- What surprised you?

- What did you learn?

- Logistics?

- What would you do differently next time?

- What do you wish you knew more about that would have helped you in your research?

# QUALITATIVE VS QUANTITATIVE DATA

## Quantitative Data

- Data can be measured and recorded in numbers

- Typically easier to measure

- Typically easier to use in comparisons

- Quantitative data is seen as more objective

## Qualitative Data

- Data is observed and recorded in words or a narrative

- Deals in descriptions and qualities

- Qualitative data is seen as more subjective

# QUALITATIVE VS QUANTITATIVE DATA COLLECTION

**Quantitative Data**

- Captured via surveys, questionnaires, and observation

- Produce more manageable results with easier analysis
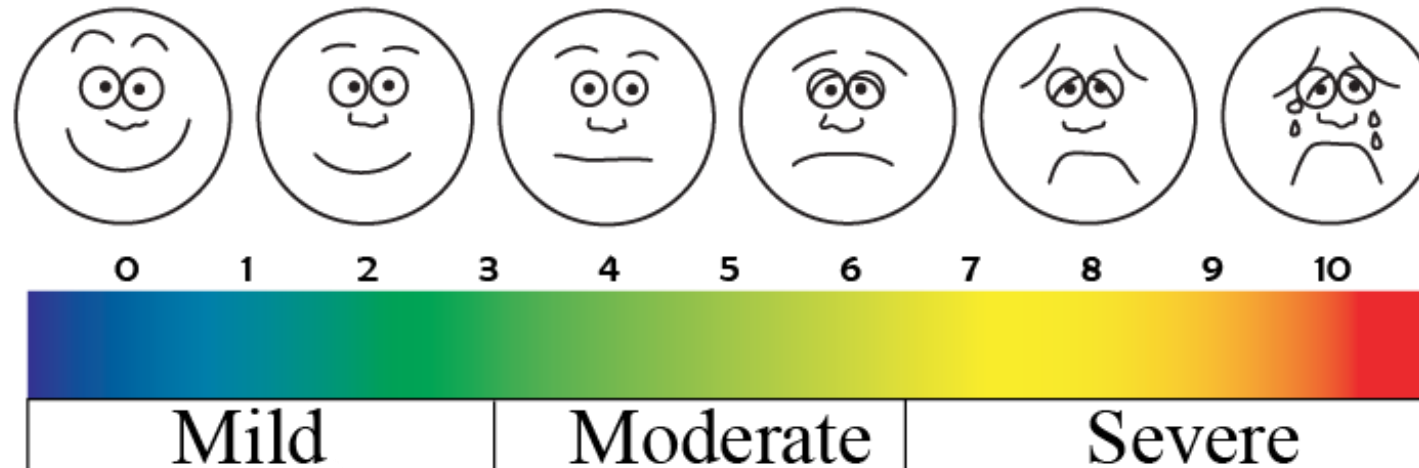
- However these methods are less flexible

**Qualitative Data**

- Captured via interviews, focus groups, content review & observation

- Tends to be harder to work with and draw conclusions on

- However these methods can capture unanticipated results.

# QUALITATIVE VS QUANTITATIVE DATA COLLECTION

Mixed methods

- Often, qualitative data is collected using a blend

# DATA ANALYSIS

**Data analysis:**

- Data is meaningless without analysis – just a bunch of words and/or numbers.

- The process of analysis aims to find themes/common threads through (sometimes disparate) data.

- Analysis can be done through quantitative or qualitative methods, or a mix of both.

# DATA ANALYSIS

**Is driven by the goal of the evaluation and the evaluation questions**

- If your goal is to prove that a certain malaria vaccine is effective – that requires different analysis (and a different level of rigor) than if your goal is to capture lessons learned from implementing public forums.

- If your goal is to produce findings that can potentially represent an entire population, that will be a different analytical framework than if you want to analyze a certain group of people.

# DATA ANALYSIS

**Audience is key:**

- If your audience is academics – your analysis will be judged in terms of rigor and contribution to theory

- If your audience is practitioners – your analysis will be judged in terms of how clear it is and how useful the findings produced are.

# DATA ANALYSIS

**There is no set formula/recipe!**

- No 'formula' for transforming data into findings.

- Texts/trainings can offer guidance, but the process is ultimately up to the evaluator!

- Evaluator has an ethical obligation to do analysis in the most rigorous manner possible, taking into account the needs of key stakeholders.

- There will always be questions about whether you picked the "best" way

# DATA ANALYSIS: JUST DIVE IN!

**You must draw a line in the sand and start doing analysis**!

• You will never have enough research or enough data.

• There will always be more data you wished you had, more questions you wished you had asked.

 - As you start analyzing your data, you will see areas where you wished you had more data! But this is the process: research…analysis….more research!

# DATA ANALYSIS IS SUBJECTIVE

- Need to acknowledge your own biases when doing analysis

- Different people can analyze the same data and come up with different findings!

  - Might see a trend you like and keep it even if it isn't supported by data

  - We all look for certain things

  - You may have an idea of what's true but then see evidence in the data that has entirely the opposite finding!

  - No coincidence that studies funded by oil companies show no evidence of climate change vs studies funded by climate change activists!

- Quality of analysis is based largely on the quality of the data collection tools.

# WAYS TO ANALYZE DATA

- Quantitative data analysis

- Qualitative data analysis

- Both quantitative and qualitative data analysis (mixed methods)

*Does anyone have any experience in these or other methods of data analysis?*

# QUALITATIVE DATA ANALYSIS

Based on qualitative data collection

- Interviews

- Focus groups

- Participant observation

- Research field notes

- Pictures

- Video

Usually done by grouping together themes amongst disparate information

# BENEFITS OF QUALITATIVE DATA ANALYSIS

- Can capture the how/why/unanticipated stories

- Can weave complex stories together

- Flexible

- Can easily present evidence in report

# DRAWBACKS OF QUALITATIVE ANALYSIS

- HUGE amount of work

- Very difficult to find trends in the data

- Relies mostly on the analyst to 'find' findings

- Very open to bias

# QUALITATIVE DATA ANALYSIS: GETTING STARTED

**When does analysis start?**

- No line dividing the end of research and beginning of analysis

- Ideas for analysis will emerge during research

- Thinking about analysis during fieldwork can sometimes cause you to do more/different fieldwork!

**When does fieldwork end?**

- Sometimes during analysis you realize that you need to do more research!

- New questions emerge that necessitate further data collection.

- On the other hand, time/resource constraints mean that you have to draw a line at some point.

- If additional fieldwork is not possible, gaps and unresolved questions are discussed in the final report for someone else (or you) to pick up at a later date!

# STEPS TO UNDERTAKING DATA ANALYSIS

**Step 1: Protect your data**

- Your data is unique and precious. No one else will be able to capture exactly what you have captured. You have a piece of a story that no one else will ever have, so you must back up your data.

  - Not enough to have it on two drives in the office – what if there is a fire?

- Back it up in a manner that protects the confidentiality of your research participants

- Protecting your data means coming up with a plan so that the data won't be accessible to people beyond the research team

# STEPS TO UNDERTAKING DATA ANALYSIS

**Step 2:  Data reduction**

- After you collect the data, you will (usually) suddenly realize how much of it you have. Can be very overwhelming!

- You need to select, focus and simplify the data, which is called "data reduction."

- Normally, evaluators do this throughout the data collection process, especially if they are collecting a lot data.

- Data reduction essentially involves findings patterns and themes. There are two ways to do this.

# DATA REDUCTION: METHOD ONE

**Inductive analysis**

- Discover patterns, themes and categories that **_emerge_** out of the data.

- Approach all your data with open eyes, no preconceived ideas of what findings you are interested in

# DATA REDUCTION: METHOD TWO

**Deductive analysis**

- Data are analyzed according to a predetermined framework.

- Come up with themes ahead of time – before you start analyzing

- Make sure stakeholders agree with themes

- Eg. Looking for evidence of motivation to vote in a data set of a 100 interview transcripts.

# INDUCTIVE OR DEDUCTIVE?

- There is no right or wrong way: develop themes first and match data to themes, or look at the data and then develop themes.

- Usually it will depend on if you are looking for specific things (the expected), or if you are unsure of what you are looking for (the unexpected).

- Sometimes, people mix it up: start with pre-determined themes, and add/subtract as they go through the data.

# CODING QUALITATIVE DATA

- <mark>This</mark> technique <mark>is</mark> used to understand qualitative data.

- <mark>An</mark> analyst assess data and codes/demarcates/identifies it

- Coding is an <mark>attempt to</mark> draw on the strengths of quantitative data analysis methods for quantitative data

- Can be applied to <mark>explain</mark> and highlight trends in collected qualitative data

- Though an imperfect method <mark>coding</mark> is an essential tool for in depth analysis of qualitative data

# CODING EXERCISE

1. Which two go together?
2. Which one is the odd one out?

# CODING DATA

**Why is this important?**

- When presented with information (data) – people will think differently.

- Often the way we see data is based on preconceived ideas we have.

- If you're working in a team, each team member might see the data differently.

- This isn't necessarily a bad thing! Viewing data from different angles can help one better understand the complexity of the data.

- But it also presents a challenge – how to objectively analyze qualitative data.

# CODING TOOLS VS. METHODS

Ultimately, coding is all about recognizing patterns (or a lack of patterns) → **the tool you use doesn't matter!**

Different kinds of coding tools:

- Special Software: Dedoose, Nvivo

- Basic Software: Word, Excel

- No Software: Highlighters, Sticky Notes

So, if your tools don't matter, what does? What makes coding different than just writing down your impression of a document?

# WHAT MAKES CODING UNIQUE?

Reducing data by…

- Reading it once & sharing your impressions = writing a summary

- Reading it more than once & sharing observed patterns = coding

- Both are subjective, but one is more methodical than the other

The biggest difference between them is **internal consistency**

When good coding is finished you should have a sense of the whole body of data

# Who makes the best tomato sauce?

# EXERCISE

- When you have read through the recipes and highlighted patterns, one easy way to organize your codes is to make a simple table in Excel.

- You don't even need to type in answers; you can just copy and paste in your highlighted codes into the right category

|  | Cooking Time | Type of Herbs | Other Vegetables | Fresh or Canned ? | Oil or Butter? |
|---|---|---|---|---|---|
| **Recipe 1** | 60 minutes | oregano | garlic | canned | oil |
| **Recipe 2** | 30 minutes | thyme | onion, garlic, carrot | canned | oil |
| **Recipe 3** | 45 minutes | none | onion | canned | butter |
| **Recipe 4** | 120 minutes | basil | garlic | fresh | oil |

# CODING IN REAL LIFE

Coding Step 1: Establishing the Purpose

- It can be tempting to just start searching through data for findings – can take a LONG time!

- It might be better to at least start with a sense of what you are interested in in the data.

  – To do so – revisit your evaluation questions

  – Based off of these questions – what sorts of things should you be looking for in the data?

  – Don't try to look for everything, particularly if you have a lot of data…the analysis will be frustrating and endless. Start by looking for things that will answer your evaluation questions!

# CODING IN REAL LIFE

Coding Step 2: Get a Feel for the Amount of Your Data, and Potential Themes.

- First, skim your data (interview transcripts, pictures, documents) and – based on your evaluation questions.

- Best to eyeball it

- As you skim, develop your initial code list

# CODING IN REAL LIFE

Coding Step 3: Code your data!

- Coding breaks data down into separate units of thought and organizes them according to themes.

- To code successfully, comb through your interview data a few times:

- Skim your transcripts first and make notes on what you can do with parts of the data.

- Develop a list of coding categories based  on patterns in data

- Review transcripts again and begin coding in a systematic way, taking parts of data and indicating how it fits within a given code (some parts of texts may fit in multiple coding categories).

- For parts of your data that don't fit – 'park it'

- Avoid changing the meaning of data when coding

# CODING DATA: EXERCISE

Example: "I wish I could have participated in the roundtable, however, I was not able to take the time off work.  Regardless, I would not have been able to get there anyway.  The event was too far away from my house and there is no public transportation I could use to get there".

**What are some potential codes?**

# CODE TYPES

**Two types of codes:**

- Objective codes:  consistent and easy to identify (everyone would code the data the same way)

    - e.g. – Party members who spoke with journalists about CISC public forum after participating in the event.

    - e.g. – Citizens who indicate that they attended the CISC public forum

- Nonobjective codes:  difficult to interpret what the data means.

    - e.g. references to learning something at the public forum events

# CODING TIPS

When coding it's important to write memos to yourself at the same time. That way you can your record your ideas, reactions, etc.

Sometimes after you code data you may find that you have to go back and collect more data if there is not enough already available. Then code some more!

Be careful to not make more of the data than what is really there! It is tempting to hang on to an interesting theme, even if it becomes clear that the theme/finding does not really bear out in the data!

# CODING TIPS

Sometimes if you see a piece of data that seems important, but you are not sure why… – you can 'park' it somewhere. Then go back and look at your parked data and see if there are other themes to explore.

Understand that by coding, you are putting your worldview on someone else's world. You are putting meanings onto someone else's perspective. One way to test the robustness of your 'worldview' is to present the codes to people whose world is being analyzed.

# WHAT DATA CAN BE CODED?

**Coding does not have to be limited to text!!!!  (photos, video, etc…)**

- Eg.  Coding newspaper articles:

- Placement of article on page

- Photo with article?

- Length of article

- Tone of article

- Theme of article

# IDENTIFYING FINDINGS THROUGH CODING?

**How do you know if you have a theme/finding in your data?**

- There is not set answer to this question.  One strategy – report how many people said x.  ie – 6 of 10 respondents indicated that they could not attend the training due to lack of transportation.

- Again, there is no hard rule here.  If you feel you have a finding, report it, but make sure you provide the data to back it up!

# VALIDITY AND RELIABILITY

# VALIDITY

Internal validity:  are there other explanations for observed results?

- Eg:  Finding, ice cream eating causes increase in spousal abuse……Alternative explanation, hotter climates cause increase in spousal abuse. In hotter climates people also eat more ice cream.  So finding was not valid.

- Eg. A certain pesticide caused an increase in crop harvest.  But perhaps also had a heavy rainfall season.

- E.g.  Training made people more motivated to undertake activism in Jordan.  But perhaps motivated people volunteered to be in training!

# HOW TO IMPROVE INTERNAL VALIDITY?

- RCT

- Random selection

- Triangulation

- Member checking

- Thick description

- Clarify researcher bias

- Prolonged field research

- Peer debriefing

- External auditor

- Consider and rule out alternative explanations

# TRIANGULATION

**What/Why:** No single method can ever answer all questions fully. Different methods reveal different facets of the truth! Triangulation uses different methods to get the whole picture.

- Uses findings from diverse sources (as opposed to findings from diverse studies, as in social science meta-analysis)

- Confirm findings, reinforce credibility

- Compare different findings

- Find inconsistencies in data

- Find opportunities for further investigation

- In M&E, triangulation is a routine activity, not a special case

# WHEN TO TRIANGULATE

Always good to triangulate, but especially:

- Insufficient or poor quality data

- Dissimilar data

- Complex questions

- When new, rigorous research is not possible

BUT:

- Takes more time

- Increases the amount of data

- Conflicts between data, investigators, methods, theories

- Interpretation and analysis complexity

# EXTERNAL VALIDITY

**External validity is the ability of the finding to be generalized to other contexts.**

- Finding: *Public forums increase the motivation of citizens to participate in politics.*

- Does this translate to other contexts? Will this work in other countries?

**How to improve external validity?**

- Try to ensure you research a random sample

- Try to stratify your sample to match the population

- Describe the context surrounding the finding

- Do study in other places to see if the finding bears out!  (replication)

# RELIABILITY

Reliability is the chance that you will get the same finding more than once

- Check transcripts

- Consistency of codes (in both their meaning and potential utilization)

- Inter-coder agreement: between at least two coders compare coding results (80%)

- Sharing analysis

- Regular meeting of analysts

- Minimize factors that can influence findings (administer survey same time of day for example)

# QUESTIONS AND ANSWERS

Qualitative Data Analysis & Coding

MOHAMMAD ALLAHHAM
REGIONAL M&E COORDINATOR
MALLAHHAM@IRI.ORG