

## Brief Comments about Qualitative Data Analysis

There are three vital aspects of qualitative data analysis. (1) What are the data “telling” you? What patterns, themes, and concepts emerge from the data? (2) Do the themes, patterns and concepts that emerge from the data conform to what my theoretical framework indicates should emerge? Am I seeing what I “expected” to see – based on theory – or are there unanticipated results? If so, what do those results mean? (3) How can I validate the conclusions that I draw from the data? Have I made sure that I have taken appropriate steps to ensure that the conclusions that I draw from the data are justified?

The following are some common components in qualitative data analysis (and, as far as that goes in statistical analyses as well). Most researchers use several of these components in the data analysis process. Many use all of them. These components overlap in most cases, and may be completed concurrently rather than sequentially. In fact, I would argue that the most fruitful approaches are iterative where the researcher continually formulates conclusions based on partial data and then goes back and tests those conclusions and develops new conclusions as the data are collected. I am making the assumption in this document that the data you have are complex, the kind of data normally collected through interviews, focus groups, or observation – not the data from checking boxes on a questionnaire.

### The Components

Data archiving

- Descriptive coding

Exploring the case or respondent

- Understanding the case (memos to self)

- Coding by topic

- Analytic coding

Finding similarities between cases

- Finding themes

- Creating categories

Higher level abstraction

- Finding relationships between categories

- Creating typologies

Explaining and understanding

- Seeking synthesis

- Understanding differences

Validating

### Data Archiving & Retrieval

Being able to retrieve all of the information from a specific case or respondent is critical for analyzing complex data. For simpler data, there is usually no need to “revisit” the original data from each respondent. Once you code and enter the data into a spreadsheet, you may never need to refer to the original data again. Most researchers “revisit” the original data from interviews, observations, focus groups and such many times during the process of data analysis. Therefore you need to be able to retrieve any entire data file easily. This is simply a process of identifying the cases or respondent in some way and then storing all of the information you got from that case or respondent in a retrievable form. This may be on paper, in a computer program, or in recorded media.

## Exploring the Individual Case or Response

Exploring and extracting all of the useful information from each case or response is a key step in qualitative data analysis. There are many ways to do this, but many researchers find three processes very useful: (1) writing memos, (2) coding by topic, and (3) analytic coding.

### **Memos**

Memos are your comments about the data – your initial attempt to understand what your data are “telling” you. Memos are useful at every step of the data process, but in my opinion the original memos are absolutely critical. This is your opportunity to really think about the data you have collected. Memos can take many forms – questions to yourself (“What did (s)he mean by that?”), identification of key phrases that catch your attention (“Great way to say that!”), comments about surprising ideas that emerged (“This is a barrier that I never considered.”), or even your initial interpretation of what the respondent’s answer implies (“Seems like this implies that government agencies are discriminatory.”). Therefore I highly recommend as a very first step in qualitative data analysis a review of each individual data record. In fact, in face to face (including polycom and such) data collection, memoing should start during the interchange itself. As you interact with a respondent during an interview, for example, you should have an instrument that will allow you to note facial expressions, body language and other non-verbal forms of communication. Similarly, with data collection methods like focus groups, part of your job as the researcher is to observe and note the interactions among group members, how group dynamics develop, and the roles of individuals within the group.

### **Coding**

Virtually all data analysis, with or without the use of statistics, requires coding. People may answer “yes” or “no” to a question on a questionnaire, for example, and you will code this as a “1” for yes and a “0” for no in a spreadsheet. However, the purposes for coding complex data are quite different than those for coding other kinds of data. Table 1 compares some key aspects of coding for simpler (check a box on a questionnaire) and complex (narrative, observation) data. There are at least two main objectives for coding data for complex data.

The first is to be able to retrieve all of the data about a specific topic. These topics are usually determined by your research questions (*what* you are exploring) and are normally reflected in your research instrument (like an interview schedule). I will use a research example here about the needs of Hispanic and Latino farmers and ranchers (H/L for short). One topic that we explored was their goals for their farm operations. In this case, you will want to be able to identify and retrieve all of the information about this **topic** across all of the cases. You can assign a code to each topic (like “primary goal” and “secondary goal”). Again, you can store the information in any format, but electronic retrieval is easiest. This can be something like a searchable database where the actual verbatim (transcription) responses of the participants are stored or you can use something like a digital tape recording and simply store the place on the tape where the topic was addressed.

The second is to understand and **start to analyze** the data. This is where you start to develop categories. Analytic coding is the first step in understanding how the cases or respondents are similar or different, in “making sense” of the information you have. Analytic coding, unlike topical coding, requires **interpreting** the data. One topic for the H/L project was goals for the farm operation. One H/L respondent might say: “I don’t really want to do much different with my farm.

We're pretty pleased with the way things are going now." Another respondent might say: "We would sure like to be able to make a living from farming." Yet another might say: "We operate like my daddy and my granddaddy did. It's part of our family tradition and we want to keep it that way." As you read or listen to these responses, you may code the first response as "maintain status quo," the second as "farm full time," and the third as "maintain family tradition." You are interpreting the data. You are taking the concrete comments that the respondents made and assigning an abstract meaning to them.

Table 1. The role of coding in data analysis

| Aspect of Coding              | Simpler Data  | Complex Data   |
|-------------------------------|---|--|
| Place in the research process | Usually occurs just once, after data collection is complete and prior to data analysis          | Usually occurs throughout the research process   |
| Why code                      | Code summarizes (as in a scalar score) or represents (0 and 1 for no and yes) the original data | Code is more like a filing system – it points you to the original data, tells you what's "interesting" about them and where to find them. The objective is both to allow you to easily retrieve and re-examine the original data by topic, and to allow you to start to "make sense" of or interpret the data. |
| Who does it                   | Usually treated as a clerical task; no special expertise needed                                 | Part of the researcher's task because the coding requires extensive knowledge of the research project, the theoretical framework for it, and the data analysis process   |

### Finding Similarities among Cases or Respondents

Much more commonly, the researcher's objective ultimately is to find the commonalities and differences among the cases or respondents. There are many ways to go about this, but I will discuss two fairly common procedures.

#### Common Themes

Many researchers first search for common themes that emerge from the data (Table 2). This often emerges from the topical and analytical coding process. Taking the H/L example and the topic of goals, we might have a series of responses like these. (1) "I don't really want to do much different with my farm. We're pretty pleased with the way things are going now." (2) "Things seem to be going pretty good. I can't think of much to do differently." (3) "We're happy with how the farm is now." In your analytic coding, you might have assigned the abstract meaning of "maintain farming operation" to all of these responses. As you look at more and more cases, you may decide that one **theme** that has emerged is that of maintaining the farm as it is. In other cases, themes emerge that are not directly related to any of your topics. For example, you might find comments like these to questions that address different topics during the interviews with H/L farmers. (1) In a question about where people get information, one respondent says: "I try to ask my neighbors for help, but they are mostly Anglos. They don't tell me much." (2) In another question about how satisfied farmers are with USDA programs, someone responds: "USDA seems to discriminate against Latinos. Their information is always

in English and I never get a call back.” (3) In a third question about local Extension programs a farmer says: “Most county Extension agents don’t like to work with Hispanic farmers.” As you note these comments, you may decide that “perception of discrimination” is a theme that has emerged.

### Categories

Ultimately, just as statistical analysis collapses individual cases or respondents into groups, most qualitative analysis has the same goal. There are some research projects where this does not occur. The “one case” study is an example. However, there are serious weaknesses with this design with regard to all forms of validity (see deVaus for a good discussion of this issue) and I will ignore that instance here. Categories may emerge based on the topical and analytic coding, but they can also emerge based on unanticipated themes or characteristics of the respondents. Using the H/L example of goal for the farm, we might eventually develop several categories of goals for the farming operation and then re-examine all of the cases to see if they “fit” into those categories. We might also re-examine the cases once the theme of “perceived discrimination” emerged and decide that farmers can be categorized by their perception of discrimination – perhaps those who see discrimination as an obstacle to their success, those who mention discrimination but do not seem to feel that it is an obstacle for them, and those who do not mention discrimination. Membership in one category (maintain farm operation) does not preclude membership in other categories (strongly affected by discrimination). Similarly, categories are typically fluid during the data analysis process and their definition and characteristics may and probably will change as more cases are studied.

Table 2: Themes and categories

|                                  | <b>Simpler Data</b>   | <b>Complex Data</b>   |
|----------------------------------|---|---|
| Themes                           | Usually predetermined “Check the box that best describes your goals for your farm or ranch.”  | Emerge from the data as the researcher interprets the respondents’ comments and assigns an abstract meaning to them. Themes may be related to topics identified by the researcher, but may also be unanticipated and not related to any specific topic.   |
| When categories are established  | Predetermined, usually at the time of developing the research design. Major changes are usually not possible because the categories are ‘built into’ the research design and instruments. However, original categories are sometimes collapsed and <i>post hoc</i> analysis can create new categories | Categories are generated during the analysis process itself. New categories emerge; some categories disappear; categories are merged into larger categories as data analysis proceeds. Developing the categories is <b>part of</b> and may be a primary objective of the data analysis process. |
| Flexible or exclusive membership | Usually exclusive membership; e.g., membership in one category only   | Membership may overlap; the same respondent or case may belong to several categories that emerge  |

## **Modeling**

Some researchers go a step further and explore the relationships between the characteristics of the respondents and/or the categories of which they are members. This is probably most similar to cluster analysis in the statistical approach, and in fact some researchers use a non-statistical form of cluster analysis to analyze their data. Essentially, you are looking for characteristics of the respondents that co-vary or “go together” across categories. At this point in your data analysis, you start to concentrate on exploring the similarities and differences between cases. My recommendation, and that of many others, is that you do this as an integral part of the data analysis process throughout the process, rather than sequentially. I often recommend that case comparisons be made very time four or five new cases have been added to the data base.

### ***Relationships between Categories***

Just as membership in a category is not usually exclusive, there can be multiple relationships between categories. Taking the H/L example, we could categorize farmers by farming experience. Perhaps we have three categories, new or beginning farmers (little previous farming experience), farmers who have considerable experience with farming, but not in the United States, and farmers with extensive experience in farming in the U.S. We have also categorized our respondents with respect to their perception of discrimination. As you compare and contrast cases you may find a relationship between farming experience and perception of discrimination. Perhaps new farmers and farmers with little experience farming in the U.S. are more likely to perceive of discrimination as a barrier to their success than farmers with extensive experience farming in this country. Discerning the relationships between categories involves higher level abstraction in which you not only assign abstract meanings to the categories themselves, but also postulate relationships between the abstract categories that you have created. This is a fundamental step for gaining an in-depth understanding of complex information.

### ***Typologies***

While not all typologies are based on categories, it is difficult to imagine creating a typology without first generating some kind of abstract sets or categories. At this higher level of abstraction, you can generate broader groupings of cases and describe their commonalities. In some ways, a typology is an exhaustive set of groups that includes all (or at least almost all) of your respondents and allows you to “paint a picture” of what each group looks like – what characterizes them. Just as membership in a category is typically not exclusive of membership in other categories, typologies may combine different kinds of categories to create more comprehensive descriptions of social groupings.

In the H/L example, we have talked about categories of farmers based on experience farming and categories of farmers based on their goals for their farm operations. We have one group or category of farmers identified called “new or beginning farmers” and another group identified by goal as “expand farm operation.” Our goal in this project was to create an action plan for USDA to more effectively address the needs and concerns of H/L farmers and ranchers. We ended up with a typology that actually combines attributes of several ways of categorizing farmers. For example, one type in our typology is “new or beginning farmers.” They are characterized primarily by their experience farming, but actually overlap in many ways with another type called “expand farm operations.” New farmers are just that – new to farming – but they share some characteristics with the “expand farm operations” type in several senses. Other characteristics are quite different for the two. First, new farmers may eventually want to expand their farm operations – but they are just getting their feet wet, trying out farming, so their ultimate goals for

their farms are apt to be less well formulated than those of the “expand farm operations” type, or at least less immediate. Both of these types are apt to be open to experimentation with adding new enterprises to their farm operations because both are trying to find “things that work” to make their farms a success. New farmers are apt to be avid information seekers – after all, they’re new to farming. The “expand farm operations” type tends to be more experienced in farming. They may be much more discerning and demanding in terms of what information they find useful. The new farmers are in an expansive “search mode” for ideas that will work. Farmers in the “expand farm operations” type have often tried and discarded many options, and they have a more focused “search horizon” for workable ideas.

In short, the objective with creating a typology is not necessarily (or perhaps even commonly) to create a set of types where every case or respondent neatly fits into one and only one type, but rather to capture the key features or most telling aspects of each type, understanding that there are usually not sharp lines between them. In one sense, you are trying to summarize and generalize what you have learned to describe “ideal types” that may not exist in their “pure” form or at least may not be very frequent as “pure examples” among the cases. However, these ideal types capture the most salient aspects of each type **and** point to the distinctive differences between them.

## **Understanding and Explaining**

For many researchers (but not all), understanding is the final goal of data analysis. The researcher wants to go beyond describing what exists to **explaining** the relationships that he/she has identified during the data analysis process. For purposes of this class, we can think about two approaches – finding synthesis and understanding differences. In reality, most research requires that you do both. In fact, you will probably use several of these approaches in any single analysis, often repeatedly. However, I think it is helpful to think through the two ways of thinking about your data. Richards (2007, p. 132-133) suggests four ways to approach each aspect of data analysis.

### ***Seeking Synthesis***

For some researchers, the goal of the data analysis is to “explain the big picture” – to synthesize. One way to think about this is that you want to “tell a story.” You might think of this as distilling all of the complexity that you discovered and thought about during data analysis into an account of what is central or crucial to understanding the phenomena you explored.

The Big Picture. Create a unified account of what the data illustrate. Focus on the central themes or relationships that emerge. Emphasize clarification. Try to understand your data from the perspective of the dominant features that emerged. Ask yourself, “Can I explain why these themes or relationships are so dominant in my data? Do I understand why these themes and relationships appeared and not others?”

The Pathway. Create an account of “what happened.” Focus on the processes that emerged in your data. Explain, for example, how one set of decisions led to another decision point, that then led to another decision point. Ask yourself, “Can I explain the outcomes, the end state, of the phenomena that I studied? Do I understand how the different cases or respondents ‘ended up’ where they did?”

The X-Ray. Create an explanation of a puzzling or poorly understood behavior or social phenomenon. Focus on the key elements (themes, relationships, patterns) in your data that help

us understand something that initially seemed “inexplicable.” Ask yourself, “Can I make sense of this phenomenon now? Do I understand *why* this happened?”

The Common Basis. Create an explanation of why a complex phenomena “works,” why patterns repeat themselves. Focus on the commonalities between what initially appear to be very different outcomes. Look at what different processes, social groupings, organizations, etc. share in common. Ask yourself: “Can I find common mechanisms that work across all of the different types I identified? Do I understand what it is that they all have in common, despite their apparent differences?”

### ***Understanding Differences***

For many researchers, understanding the differences or how cases diverge can be more important and more telling than understanding what they have in common. The researcher wants to make sense of diversity, to focus on the contrasts, not the commonalities.

The Dominant Pattern. Create an explanation of why some cases differ from the dominant theme. This is often a useful counterpoint to synthesis. After you have identified the commonalities, the dominant themes, patterns or processes, re-examine the cases. Find the outliers. Ask yourself: “What made these cases different? Do I understand what specific attributes of these cases overrode the dominant themes, patterns or processes?”

The Animated Model. Create several models to explain the different patterns, themes and processes that emerge from the data. Focus on complexity and multiple models, not simplicity and synthetic models. This is often a useful step in arriving at synthesis because you can use your multiple models to find what they have in common. Ask yourself: “Can I propose a model for each type or category that I have identified? Do I understand what the models have in common and where they differ?”

The Classic Cases. Select “classic” cases from your database that exemplify the key themes, processes or patterns that have emerged from your data. Focus on the cases that best exemplify the dominant features of the data that you want to explore. Focus on cases that “grab your attention.” Ask yourself: “What makes this case so interesting? What aspect of what I have discovered does it illustrate?”

The Zoom Lens. Put an existing theory to the test. See if it does, in fact, explain or allow you to understand what you actually found in your cases. Focus on whether the theory is only a good “general fit,” or whether it actually allows you to understand and interpret the intricacies of the real life cases that you explored. Ask yourself: “How useful was my theory in helping me understand what I actually learned about these cases? If it failed to provide a complete understanding, how could I expand or elaborate the theory so that it is more useful?”

### **Theory Building – Theory Testing**

I personally think that the distinction between theory building and theory testing is somewhat overblown. I would argue that most of us are actually doing both things most of the time. However we may choose to analyze data, it is the researcher’s responsibility to both contribute to theoretical understanding **and** to examine how the data challenge existing theoretical frameworks. Therefore, I would argue that it is more a question of where we start than where we end, that all theory grows out of observations (is grounded in what we observe) and that all observation should be undertaken with careful attention paid to challenging existing theories or

understandings. Certainly, qualitative data analysis requires both constructing understanding or explanation **and simultaneously testing** those understandings or explanations. Otherwise, we run the risk of simply confirming our own understanding because we failed to consider other explanations or understandings of what we observe. To try to simplify this somewhat, I suggest below some ideas that may help in testing your ideas – trying to make sure that your conclusions are justified. Again, I would encourage you to use an iterative approach and not strictly divide your analysis into two distinct tasks (build versus test), nor think about this as a linear process, whether you use a qualitative or a statistical approach to data analysis. I would also argue that even if your primary goal is to build theory through a grounded approach, you also have a responsibility to continually test your own conclusions. E.g., the theory you are testing may be your own!

### ***Outliers***

Pay careful attention to outliers at every step. This can be something like a phrase or response that seems out of place or unrelated to what you were discussing in an interview, cases that will not neatly fit into any of your categories of types, or themes that emerge rarely. Outliers can be important indications that you are oversimplifying. If those themes, patterns or relationships that “don’t fit” keep piling up, you need to re-examine how you are identifying commonalities. Even the rare outlier may be alerting you to unexplored areas or, in some cases, to whole missing segments of the range of responses that exist in the population. That one person who said something “weird” may actually represent an entire group of respondents that you failed to include in your sample. Don’t disregard outliers. Examine them carefully.

### ***You Only Hear It the First Time Once***

I personally believe that your initial response to the data is very important, particularly data from interviews, observations and such. What we see as unique and interesting quickly becomes the norm. I therefore highly encourage you to avoid over-reliance on recordings and other forms of “stored” data. Pay careful attention the first time you look at or hear the data. Make copious notes about what catches your attention. If the things you hear in the early interviews become “commonplace” and no longer catch your attention, that’s fine. Perhaps these really are common themes or ideas, but keep listening so that you can hear the new messages as data collection proceeds.

### ***Analyze Early and Often***

I would argue that saving all of the data to hear or see at one time is a mistake in most cases. As themes and ideas emerge, you need to be able to pursue those with new cases or respondents. Part of the beauty of complex data is that you get ideas **from** the data themselves. Whole new areas of questioning or interest can emerge from what you learn. You can’t make much use of that learning experience if you wait to analyze the data until the data collection process is complete. I think it is important to review every case or response as soon as possible after you first collect the data and to frequently compare and analyze across cases so that you can take full advantage of your emerging understanding to improve the quality of the data you collect. Incorporating themes or ideas that you did **not** include in your original protocol is one good way to make sure that you don’t simply confirm what you already thought was true.



### ***Look for Alternatives***

Actively seek out (in the literature) and examine alternative explanations to your own emerging ideas or theoretical framework. Purposefully include rival ways of organizing the data (categories, themes, etc.) and rival explanations into your theory building and testing. Keep asking yourself “Do I have the best possible way of organizing and understanding my data?”

### ***The Devil Is in the Details***

As you develop your understanding and try to identify dominant themes, patterns and processes, make sure you are not over-generalizing. Synthesis is great. But synthesis can also mean glossing over differences that really are important. Apply your abstractions and understanding to the most complex, the most difficult to understand cases or responses. Try to establish a balance between generalization and depth of explanation. Focus on what your model, typology, or explanation **doesn't** explain.

### **Validity & Reliability**

All of the procedures and “self checks” described above help make sure that you reach valid and reliable conclusions. Refer to the readings from week 4 of this class – Overview of Basic Research Design Groups. There were some excellent materials about reliability and validity in qualitative data analysis. As a sort of final check, I would strongly encourage you to use the following techniques – all of which were described in more detail (and better) in the readings from that week.

1. Keep a good record (or log trail) of everything you do. Always, at every step, write down exactly what you did. This helps you remember – and helps keep you from skipping important parts of the data collection and analysis process.
2. Be consistent and transparent. This is critical for reliability. Any of us would obviously find it difficult to accept the findings from a study if the researcher did not use consistent ways of getting, processing and analyzing the data. Consistent does **not** mean that you have to have a standardized interview protocol and always ask every participant the same questions, for example. It does mean that you have to have a consistent set of guidelines that you use to guide your data collection and analysis and that you justify your decisions – which may include deciding to add some questions about a new topic, to use probing questions, etc.
3. If it's important information, get it at least twice. This is sometimes called triangulation, although I have seen that taken all too literally to mean “ask it three times.” The idea is more subtle and involves building in cross-checking into your data collection. This is partly because you may “miss one” in the data collection process – e.g., your question 5 was poorly worded and people didn't understand it. Perhaps even more important, asking about an important theme or issue in a couple of ways, or at a couple of points during an interview, helps ensure that you are “hearing” and interpreting the response consistently.
4. Response checking involves giving the person who responded to you an opportunity to make sure that you interpreted what he/she said or did correctly. One way is to simply give the individual a summarized version of what you thought he/she said. We did this on the H/L study. This is not a verbatim transcript of an interview or anything nearly as time-consuming for the respondent. It's a summary of the key points that you thought the person tried to convey.

5. Member checking refers to going back to respondents (all or some of them) with your ideas at or near the end of the data analysis process. Basically, this is a way of making sure that your *interpretation* of the data makes sense to the people involved in the study. There is, of course, a potential problem with this approach. Your understanding, even if it is different from that of the participants, may actually be better. All of us routinely misunderstand our own behavior, motivations, and decisions – and sometimes just plain fool ourselves about what we are doing. I am not a partisan of the idea that I'm an expert on a subject just because it happened to me. Battered women, for example, may have a great understanding of what it means to be battered – but very little understanding of why they remain in abusive relationships. In fact, I would argue that you have a responsibility as a researcher and expert in the subject under study to have a more profound understanding of the phenomena than lay people. So you have to use this with care. Perhaps it is better to think of this as a “reality check” than a “validity check” in many cases. Put another way, be careful about what you ask people to verify. Make sure you ask them to verify things that they *can* verify. And be careful not to discard accumulated knowledge and theory too quickly.

6. Coder verification. There are statistical ways of verifying that any coder or group of coders is consistent. These are often not very helpful in qualitative data analysis, however. I would suggest that you get colleagues to verify your interpretations at various points. For example, in the H/L project example, you could get some of your colleagues or your advisor to look at some of the passages and how you coded them. See if they agree with you. It is also very useful to get feedback from colleagues as you develop categories, try to identify dominant themes, and propose explanations and understandings of the data. In short, without taking up too much time, try to get others who share your research interests to examine your thought processes and help you pose alternatives. This “goading” to get you to think about alternatives is probably more important than simple verification that you are being consistent and logical.