# 2

# PLANNING AND PREPARING THE ANALYSIS

**LEARNING OBJECTIVES**

After reading this chapter, you should know how to:

- Establish an analytic objective
- Develop an analysis plan
- Choose an appropriate analytic approach
- Define appropriate boundaries for the analysis

When working with text data there is one essential, basic analytic technique: reading the text. A targeted, goal-driven analysis requires additional steps. The larger the analytic task, the greater the amount of text to be analyzed, or the more analysts whose work must be coordinated, the more important it is to develop an analysis plan before the text is read and preferably before the data are collected. A good number of books describe how to conduct a particular type of qualitative data analysis, but very little attention is given in the literature to the planning phase. How are analytic objectives established, and how do they relate to the analysis? How do you bound your data set? What is the primary purpose of the analysis? And who will be the primary audience for, and ultimate judge of, your analysis? The answers to these questions will affect your selection of an effective and efficient analysis strategy. In the pages that follow, we address these and other questions pertinent to analysis planning and provide some practical suggestions and tools for conducting an analysis that will best fit a given context.

## ESTABLISHING ANALYTIC OBJECTIVES

The first step in developing an effective analysis plan is to establish clear analytic objectives. The approach taken in developing the plan will be somewhat determined by whether the data will be analyzed in real time, as they are generated, or if you are developing an analysis plan after the data have been collected, processed, and cleaned. In classic qualitative research, at least some of the data are analyzed as they are collected, and the results may be used in an iterative fashion to modify the data collection itself.

For example, if the purpose of the research is to explore a little-known or poorly understood phenomenon, the early stages of research often entail learning what questions to ask, how to ask them, and who to ask them of. This type of exploratory interview is described in detail in Schensul, Schensul, and LeCompte (1999). The approach can be fluid and dynamic, with the interviewer(s) making adjustments to the interview process in real time. Or, the approach may be purposely staggered, with interview guides being semistructured during each phase but the content of the guides being driven by what is learned from each successive phase. A more limited structured iterative approach may be used in situations where considerable knowledge exists about the research topic or phenomenon but the work is being conducted in a new context. In this case, the iterative approach may be limited to piloting of a more structured data collection strategy to ensure that the recruitment process, wording of questions, and ability to uncover new dimensions are effective for the new context. In each of these situations, on-going data analysis is crucial to the success of the endeavor.

Since our focus in this book is on data analysis, we will not get into the nuances of iterative data collection except to discuss the analytic implications of a dynamic and evolving database. In addition, our discussion assumes that more than one person is collecting the data, that is, that a data collection team or field team exists. Where appropriate, alternative strategies for one-person analysis projects are provided.

### Analysis to Improve the Data Collection Process

As noted, the initial steps in the analysis process may be used to improve data collection when analysis begins before data collection is complete. To achieve this objective, an initial analysis plan should be developed with a focus on two objectives.

First, the analysis should serve to *enhance the overall quality of the data*. Are the data collectors following the protocol and asking the intended questions? Are they probing appropriately to generate richness and depth from study participants? Are descriptive notes included with an appropriate level of detail? If the data are being translated, are the appropriate nuances being captured? Is there evidence of rapport between the data collectors and the people participating in the research?

If you look closely at each of these questions, the necessary components of an analysis plan to enhance overall data quality become evident: Assess the consistency of questions asked by interviewers and evaluate the skills of the interviewers (technical and interpersonal). For example, to determine whether data collectors are asking intended questions, you could generate a report that looks something like Table 2.1 by listening to the audio from each interview or focus group and keeping a running tally of the questions asked. It can also be generated by reviewing transcripts from the interviews or focus groups. This approach works well for studies that use a more structured approach for qualitative data collection.

If data collection is less structured, or if concerns exist that data collectors are deviating too far from a semistructured guide, then a more nuanced approach is needed, that is, one that tracks the actual prompts used by the data collector. For this purpose, the report could include transcription of the exact wording used by the data collectors when asking each of the research questions or querying about each of the research domains. Since qualitative research places a premium on a conversational approach that allows for variability in the exact wording of questions and prompts, it is important to pay attention to that variability and assess its implications for the comparability of the data across data collection events. It is not just the responses to the intended questions that need to be analyzed, but rather the conversations that take place.

**Table 2.1** An Example of Analyzing Data in Real Time to Determine Whether Interviewers Are Asking the Intended Questions

| *Data Collector* | *Total # Interviews Conducted to Date* | *Research Question* | *Total # Interviews Where Research Question Asked* | *Proportion of Interviews Where Research Question Was Asked* |
|---|---|---|---|---|
| John | 8 | 1A: Please describe your childhood home. | 7 | 87% |
| | | 1B: What is your first memory about your childhood? | 4 | 50% |
| Mary | 5 | 1A: Please describe your childhood home. | 5 | 100% |
| | | 1B: What is your first memory about your childhood? | 2 | 40% |

Effective probing is the lynchpin of the two most commonly used qualitative data collection strategies: the open-ended interview and focus groups. To determine whether data collectors are probing appropriately, you can read transcripts or listen to recorded interviews and make note of both good and bad instances of probing. Putting that information into a summary table can also be an excellent tool for retraining to improve the skills of the data collectors. Table 2.2 is an example from a series of pilot interviews for a study conducted by MacQueen in the southern African country of Lesotho. For most studies, creating a summary table like this can be done manually, for example by copying and pasting text directly from transcripts or by transcribing short illustrative sections. With a little more effort and the appropriate computer resources, it would also be possible to add audio clips from digitally recorded interviews or focus groups so that the team can hear the conversation.

If the project is large and complex, for example, with multiple data collection teams working in multiple sites, it is helpful to develop a quality-control codebook in a qualitative data-analysis software program and then code the interviews as they come in (see Chapter 3 for a detailed discussion of codebook development). Such a codebook would include a set of codes and definitions that are based on the kinds of problems described in Table 2.2. Taking this extra step would make it possible to use the reporting functions of the software program to evaluate overall quality by site, by data collector, and by research question as well as to look for improvement over time. The resulting documentation can be very useful for deciding whether a large, complex data base is adequate for secondary analyses or subanalyses in the future, when institutional and personal memory about data collection have faded.

**Table 2.2** An Example of Analyzing Data in Real Time to Identify Problems With Probing and Develop Strategies to Improve Interviewer Skills

| ***Problem*** | ***Example*** | ***Potential Solutions*** |
|---|---|---|
| Not building rapport adequately from beginning | *Interview guide question:*<br>First, can you tell me a little about yourself? **[Let participant answer, then ask specific probes on education, etc., ONLY if not already addressed by participant.]**<br><br>*Transcript:*<br>INTERVIEWER:<br>Ok. Eh... Well I would like to know a little bit, well, about you. You see eh mmm…tell me, where did you go to school? | The opening questions should be "soft," inviting, open-ended. Let the participant tell his or her story; encourage a story. Avoid jumping into closed-ended questions. |

**Table 2.2** (Continued)

| *Problem* | *Example* | *Potential Solutions* |
|---|---|---|
| Failure to probe | INTERVIEWER:<br>Mmhm, what qualities do you look for in a person to end up saying you love him?<br><br>PARTICIPANT:<br>I look at how genuine his love is.<br><br>INTERVIEWER:<br>Mmhm<br><br>PARTICIPANT:<br>Yes<br><br>INTERVIEWER:<br>Ok, … | What does the participant mean when she says "I look at how genuine his love is"? |
| Completing statements for the participant | PARTICIPANT:<br>Even if I don't have money to go to the doctor ...<br><br>INTERVIEWER:<br>Yes.<br><br>PARTICIPANT:<br>She ends up giving me money so that I see to it that—<br><br>INTERVIEWER:<br>You go see the doctor.<br><br>PARTICIPANT:<br>I see the doctor.<br><br>INTERVIEWER:<br>Yes.<br><br>PARTICIPANT:<br>Yes. | Let the participant complete the statement. He might have completed the comment ("she ends up giving me money so that I see to it that—") in a different way, or offered more information. |

*(Continued)*

**Table 2.2** (Continued)

| *Problem* | *Example* | *Potential Solutions* |
|---|---|---|
| Completing statements for the participant | INTERVIEWER:<br>Do you have a wife or a sweetheart?<br><br>PARTICIPANT:<br>Well, I may say I have a—<br><br>INTERVIEWER:<br>A girlfriend.<br><br>PARTICIPANT:<br>A certain girlfriend.<br><br>INTERVIEWER:<br>Ok, so you have girlfriend?<br><br>PARTICIPANT:<br>Yes. | Let the participant complete the sentence; he might have said something other than "girlfriend." |
| "Pushing" for a response rather than listening and thoughtfully probing | INTERVIEWER:<br>Can you explain for me the work your partner or sweetheart is doing?<br><br>PARTICIPANT:<br>Yes, well I don't like explaining that to you.<br><br>INTERVIEWER:<br>You can explain to me.<br><br>PARTICIPANT:<br>Mmm!<br><br>INTERVIEWER:<br>What kind of job? What kind of job is it that cannot be described?<br><br>PARTICIPANT:<br>Well, I don't think it's the kind of work that can be explained, truly speaking.<br><br>INTERVIEWER:<br>Ok, you are not able to explain it. | Interviewer could have asked why he did not like "explaining that"; for example, "Why is that?" instead of pushing directly for a response. |

The second objective of analysis to improve the data collection process is to *determine whether the topical content of the research should be expanded, contracted, or refined*. The goal is to assess the effectiveness of the questions in eliciting desired information. To determine whether too many questions are being asked in one interview or focus group, a report similar to Table 2.1 could be created but sorted by question rather than data collector. It is then easy to determine

whether any patterns exist with regard to missing questions. For example, if questions at the end of the interview or at the end of a section on a discrete topic tend to be missing, this may indicate fatigue on the part of either the interviewer or participant. If certain questions in the middle of an interview tend to be missing, that may indicate poor flow in the overall design of the interview guide.

This type of summary table points toward potential problems; to confirm the nature of the problem, you will need to read the transcripts and/or listen to audio of the data collection event. Are the questions too generic or so poorly specified that they fail to engage the study participants? Even if the data collectors are diligently asking the questions in the guide and making a sincere effort at probing, the resulting data may not be useful or informative. Data collectors can often identify a "problem" question within the first few interviews or focus groups and can confirm (or serve as the starting point) for this type of analysis. A quick way to track the overall quality of specific questions or topic areas is to create a report that includes the question or topic area together with the participant response from each interview. You can then look for patterns in the way the questions are asked as well as the kinds of responses that are being generated. Are the participants struggling to make sense of the question? Is the question generating information on the topic of interest? Are some of the participants interpreting the question in unexpected ways? What kind of probing is most successful at generating rich responses? In Chapter 3, we will describe a process called *structural coding* that can be implemented in most qualitative data analysis software programs that makes this kind of analysis quick and easy.

## Analysis to Answer a Research Question

Answering a research question is a more challenging analysis objective than assuring the quality of the data to be analyzed. Establishing an achievable research analysis objective requires a good match between the *view* you want to generate, the *quality of the data* available to generate that view, and the *resources and time available*.

*The view:* The analysis objective is a short description of what you want to achieve through the lens of the data. Consider Google Earth as a metaphor: Do you want to describe the broad sweep of continents and oceans, describe major features en route from start to finish, or do you want to explore the shrubs and cracked tarmac? We will refer to this aspect of the analysis objective as the view. Defining the view is a crucial first step for applied analysis, as it will bring structure to subsequent decisions in the analysis process. For example, if you want a detailed view, then you will need a codebook that gives you the conceptual viewing power of a microscope. If you want a wider view, you need the conceptual viewing power of binoculars or a telescope, or perhaps a scope mounted on a satellite. These are very different levels of thematic identification that result in very different codebooks.

*The quality of the data:* There needs to be a fit between the view you want to generate (the analysis objective) and the data at hand. This requires a scan of the

data. If your database is relatively small, for example, on a par with the length and richness of a good novel, you can simply read the text and make notes along the way. If the database is more like a trilogy or a library shelf of books of variable quality and form, then you probably want to do a more structured scan of the data. For this, many of the strategies outlined in the previous section for evaluating the quality of data as they are collected can be used.

If you have high-resolution, street-view data from all locations, unlimited resources, and no time constraints, you can, eventually, generate a global view by painstakingly connecting all the data points and systematically stepping back to identify larger and larger features in the landscape. The reverse is not true, however. You cannot drill down to a street-level view if you do not have comprehensive street-level data.

If you have fine-grained, rich information from some data collection events but not all, you may not be able to generate a coherent view at certain levels of specificity. For example, if qualitative data have significant gaps, you cannot make major recommendations for policy or action; the data may identify some of the key points or landmarks for more refined future investigation, but they are not sufficiently robust to support a comprehensive plan for moving forward.

*The resources and time available:* Equally important is an assessment of the time and resources available to conduct the analysis. If your data are sufficiently rich and detailed to permit building up a comprehensive map from street view to global, this would support a research objective centered on generating explanatory theory from empirical data, along the lines of a grounded theory approach. Conducting such an analysis, however, requires intensive effort. Many people claim to be using a grounded theory approach in their analysis, but fully developed applications of the approach are relatively rare. Grounded theory requires a painstaking, line-by-line reading of qualitative data where each statement is systematically compared and contrasted. Most people who claim to use a grounded theory approach do not analyze the data at this level of detail because they lack the time and resources or they lack data of sufficient richness to warrant such a detailed level of analysis—or both. Applied thematic analysis is an approach that explicitly takes into account the issues of resources and time as well as the quality of the data in specifying an analytic research objective. How many people will work on the analysis? What level of skill and expertise do they have? How much time will they devote to the analysis? What is the timeline for completing the analysis? What technology is available to support the analysis?

If you have rich data but limited resources and a need to generate useful results in a timely way, it can feel at times that you are throwing away valuable data. This quandary leads some qualitative researchers to attempt a detailed analysis of all the data available. Such an attempt is often counterproductive: You may end up analyzing an arbitrary subset of the data because you run out of time, resources, and energy; you may sacrifice quality checks, leading to a false sense of surety about the results; or you may spend time and money analyzing a large data set only to realize that only a small portion is relevant to the research objectives.

In applied thematic analysis, we deal with this quandary by maintaining a systematic analysis process aligned with a targeted analysis objective. You do not sacrifice quality, nor do you discard valuable data. If there are more data than you can analyze with the given resources and time frame available, it is not discarded; it is systematically cataloged so you and others can come back to it as opportunity allows. Because it is systematically cataloged, you can be strategic in your efforts to identify additional resources for analysis, whether through funding applications, internship opportunities, and so on. As with any large, complicated effort, the goal is to break it into manageable components and then systematically do the work.

You may accomplish the task in several ways. The most obvious approach is to return to the original research design that structured the data collection. If the original design included more than one research objective, then you should translate each of those objectives into an analysis objective and then outline the steps necessary to achieve that objective. You will likely find that most or all of the analysis objectives require a common set of steps. For any given objective, you may not need to go through those steps for all of the data. It may be more efficient to prioritize their completion for the full database than to repeatedly go through the same series of steps for each analysis.

For example, you generally need to catalog the type of data available, the overall quality of the data, missing components (e.g., questions that were not asked) from items in the database and other information reflective of the quality of the data. It is easier to compile that information all at once than to repeatedly go back through a large database to compile it piecemeal. For some analyses, you may be interested in knowing how frequently certain terms or phrases appear in the data. A potentially quick way to accomplish this is to run the transcribed text through an indexing program and generate a word list and frequency count. Although this sounds simple, it in fact requires some effort to refine or generate lists to omit some words (e.g., English particles such as the, a, that) and to consolidate variants (e.g., a lot, lots). Large volumes of text can take a surprisingly long time to process in this way.

If you are working with transcripts, it may also be important to process the words spoken by the study participants separately from those spoken by the data collectors. For example if one structured interview guide was used to conduct 25 interviews, then key terms embedded in the questions will automatically occur at least 25 times because the data collectors are asking the same questions in each interview. These types of considerations need to be kept in mind whenever you consider using automated procedures to scan qualitative data. We expand on word searches as an analytic technique in Chapter 5.

Another important consideration in developing an analysis plan is the way in which the results will be disseminated. Will your findings be presented as a report of recommendations to guide policy, a comprehensive dissertation to meet the requirements for an advanced academic degree, a peer-reviewed journal article to illuminate a focused topic, a book that explores multiple dimensions of a complex topic or phenomenon, a descriptive piece in the popular media, or an “in-house” report for planning purposes? A single analysis may need to meet several of these

dissemination needs simultaneously. For greatest efficiency, the plan should be sufficiently fine-grained to meet the needs of the most demanding dissemination goal. This likely means that some (but not necessarily all) aspects of the analysis are rigorous enough for peer-review publication and/or development of policy recommendations.

A further consideration in the development of an analysis plan is one directly related to the objectives of the research design. Are you seeking to explore or describe something? Do want to go a step further and explain it? Is your goal one of confirming findings from other research? Do you plan to compare data between groups? Or, are you doing a bit of each of these? Each determines what you look for and how you look for it, that is, how you code the data.

## The Quick and Targeted Analysis

Although we typically encourage a thorough and systematic analysis, we also recognize that this is not always warranted, or possible. In client-driven research, results are often expected almost as soon as data collection is complete. It is not uncommon, for example, for market researchers to write a summary report of focus group findings the day after focus groups have been completed. In this case, analysis is constrained by the client's timeline; an in-depth, systematic analysis, including transcription of the audio recording, is often not possible. The analysis is directed toward discovering high-level themes that have meaningful and practical implications.

Even in situations where time is not a factor, an exhaustive analysis may not be worth the time and effort. In certain mixed methods designs, for example, the primary purpose of collecting qualitative data can be simply to inform a quantitative instrument (more about mixed methods analyses in Chapter 8). In such a context, the analysis is narrowly targeted to inform specific response categories, question stems, or domains of inquiry on an instrument. A complete textual analysis in this case is not necessary. Formal codebooks, systematic coding of all that was said, and even transcription of data are overkill if the only purpose is to help in the design of a subsequent instrument or other research element, such as a sampling strategy. Instead of using more in-depth and time-consuming (albeit rigorous) analytic procedures, the analysis process can be streamlined by using a debriefing template, similar to what Miles and Huberman (1994) call a "contact summary sheet" (p. 51). Data collectors are instructed to debrief immediately after each data collection event. For in-depth interviews, this may be a solitary enterprise; for focus groups, it typically involves a discussion between the moderator and note-taker. As part of the debriefing exercise, data collectors fill in a form that has been created for the particular objectives associated with the data collection activities. Although the content of the form will vary according to the objectives of each individual project, we have found from our experience that requesting at least the following information is useful, both for designing subsequent instruments or research questions and for improving data collection quality.

- Basic data about the data collection event: date and location, participant type, name of data collector(s), number of people in group (focus groups only)
- Main themes that emerged
- Information that was confusing or contradictory
- Emergent questions or domains of inquiry that should be added to the subsequent instrument
- Response categories for questions
- Suggestions for improving the data collection event (techniques, questions, etc.)

## Summing Up

As we alluded to in the opening paragraph of this chapter, qualitative data analysis should be a thoughtful enterprise, not an ad hoc process. It is true that thematic analysis should be flexible and responsive to the naturally emergent nature of the process, but many factors need to be considered beforehand to ensure that your analysis is both efficient and meaningful. In Table 2.3 below, we outline some of the questions to think about before embarking on an analysis. In fact, we strongly suggest that these be well thought out during the research design stage and incorporated into the research plan/protocol.

**Table 2.3** Questions to Consider Before an Analysis

| *Question* | *Suggestions and Tips* |
|---|---|
| **What is the practical purpose of the analysis?** | Find solutions, build theory, develop an intervention, evaluate something, inform subsequent data collection, inform ongoing data collection. |
| **What is the analytic purpose?** | To: identify, explore, explain, compare, confirm, or some combination of these purposes. |
| | **Tip**—The analytic purpose should synchronize with your research objectives. |
| **How is the analysis connected to the research question(s)?** | Your analysis should directly inform one or more of your research questions. |
| | **Tip**—Before embarking on analysis, the research team should review the study's research questions and objectives to refresh their focus and make sure the analysis is framed to inform these. |
| **What is my timeline?** | Ask yourself how fast findings from your analysis are needed. The answer can range from "tomorrow" to "no foreseeable deadline." |
| | **Tip**—If an immediate turnaround is needed, you may have to forego transcription and/or a systematic analysis. Use a debriefing template to expedite analysis. |

*(Continued)*

**Table 2.3** (Continued)

| *Question* | *Suggestions and Tips* |
|---|---|
| | **Tip**—If you have a relatively distant timeline, prioritize which analyses are most important and execute accordingly. Identify target dates and milestones to keep the analysis on track. |
| **What resources do I have at my disposal?** | You may be the sole researcher on a project, or there may be multiple analysts. You may have access to only certain analysis software or no access at all. |
| | **Tip**—Having more analysts contributing to an analysis can speed up the process, provided they have adequate access to computers and software. However, the issue of coding reliability needs to be addressed (see Chapter 4). In general, working in teams requires more quality control checks (for a guide to doing qualitative research in teams, see Guest & MacQueen, 2008). |
| | **Tip**—Select analysis software that best facilitates your analytic needs. We provide information on software in Chapter 9. |
| **How large is my data set?** | The size of a qualitative data set can range from a few in-depth interviews or focus groups to hundreds of various types of data collection activities (observation interview, focus groups, secondary data). |
| | **Tip**—Trying to include all data from all sources in a large study is cumbersome and usually not necessary (though tempting). If you have a large data set, try to divide it up into separate analyses. An easy rule of thumb is to equate a specific analysis with a specific output, such as a list of response categories, a report for a funder, or a peer-reviewed article. |
| | **Tip**—Prioritize your analyses so that the most important and time-sensitive analyses are conducted first. Consider also how analyses can most efficiently build on each other. |
| **How heterogeneous are my data types?** | You may have data from only one type of activity, such as in-depth interviews, but in a larger study, you may have data from focus groups, observation, and secondary sources as well. You may also have quantitative data relevant to your analytic objectives. |

| *Question* | *Suggestions and Tips* |
|---|---|
| | **Tip**—Think about how you will integrate data from different types of data collection methods. Decide whether data will be pooled or analyzed separately. Will you use the same codebook for two or more types of data? Answers to these questions depend on how similar and how structured data collection instruments are, sampling methods for each data type, and your research objectives. |
| | **Tip**—Chapter 8 provides more details on how to integrate qualitative and quantitative data. |
| **Which data should I use for a particular analysis?** | The answer to the question can range from a very small section of text to an entire data set, and depends on the overall size of the data set, one's research objectives, and time constraints. |
| | **Tip**—Think about which data are essential to a specific analysis. Which sources (participant or event types) of data are needed? Are responses to all questions or topics in an interview needed? This is what we refer to as "bounding the analysis." |
| | **Tip**—Use only the data that you need. Most audiences prefer a concise, nonconvoluted story, and using only the most pertinent data keeps your story line on track. |
| | **Tip**—Be even more frugal with data selection if you have a short time frame. |
| **Who is the audience for my analysis, and how will members judge the process and subsequent findings?** | You may only be concerned with writing a report for one audience, or you may need to write for several different audiences. Think carefully about for whom a particular analysis is intended. |
| | **Tip**—Different audiences will likely require different levels of rigor. Peer-reviewed journals generally have higher standards and focus on methods. Note also that expectations vary from journal to journal and discipline to discipline. Smaller sample sizes, for example, are more acceptable in anthropology journals than in public health journals. |
| | **Tip**—If possible, choose your audience based on the type of analyses your data will support. |

*(Continued)*

**Table 2.3** (Continued)

| *Question* | *Suggestions and Tips* |
|---|---|
| | **Tip**—Ask yourself, "What type of data output will my audience want to see?" Some audiences prefer narratives with verbatim quotes interspersed throughout. Others may be more amenable to matrices or tables. Still others may have a predilection for numbers, in which case data-reduction techniques will likely be used (see Chapter 6). Whatever the output, make sure that your analysis is set up to get you there. |

## WRITING AN ANALYSIS PLAN

Once you have a good grasp of the desired view, the quality of the data, and the resources and time available, you are ready to develop an analysis plan. As with many writing tasks, a good place to begin is with an outline of the major decisions that need to be made.

First, specify the kind of report or manuscript the analysis is intended to support. A single study may lead to a single report, especially in the applied research context. However, many studies are sufficiently rich to generate multiple analysis projects. This may include multiple analyses for discrete sections of a report, chapters in a book, or targeted articles for specific peer-reviewed journals. If the study will require multiple analysis projects, it is good to briefly describe each of the planned analyses, though you may start by developing detailed analysis plans for only one or two. Knowing that the other analyses are in the queue can help to keep the first analysis project bounded and achievable.

Once a specific analysis objective has been identified and briefly described, you will need to decide on the basic analytic approach to be used to achieve the objective: exploratory, explanatory, or confirmatory. Each of these approaches is described in more detail below. Next you need to determine the specific data that will be used for the analysis and why. This decision, which we describe as *bounding the view,* is described in more detail below.

Once the boundaries of the analysis have been defined, you need to decide the specifics of how the text data will be coded. There are three components to this decision. First, you need to decide what tools you are going to use in the coding process. Choosing a qualitative data analysis (QDA) software program is an obvious first step. Depending on the capabilities of the QDA program and the complexity of the data, you may also need to choose programs to assist with tracking the analysis process. For example, for a large, complex study you may

want to use a spreadsheet or quantitative database program to track specific analysis tasks such as coding assignments and intercoder agreement assessments. Second, you need to specify the codebook development process, especially if there will be more than one coder. Will there be specific formats used for codes (e.g., prefixes, numbering, standardized abbreviations) or will coding be done in vivo (i.e., coding for words or phrases within the text)? Will you use a structured codebook definition (e.g., with explicit instructions on when to use and when not to use the code), or will code definitions be developed along the lines of coding notes that are associated with specific text? Third, if there is more than one coder working with the data, you will need to specify the steps taken to ensure that the coders agree about the code definitions and which codes appropriately describe the meaning in the text. In Table 2.4, we provide a list of suggestions of items to include, or at least consider, in a qualitative analysis plan.

**Table 2.4** Items to Consider for Inclusion in an Analysis Plan

- Specify how many separate analyses will be conducted and the timeline for each
- For each separate analysis specify:
  - Which research question(s) it will inform and how
  - Precisely which data will be used
  - How many people will be involved in the analysis and their specific roles
  - The primary analytic purpose—e.g., to identify, explore, explain, confirm, compare (note the verb used is very important so choose carefully)
  - How codes will be created and defined, including structural codes
  - Rules for applying codes to the data (e.g., will all text be coded?)
  - How coding reliability will be established, including reconciling discrepancies
  - Which data reduction techniques, if any, will be applied
  - Which between-group comparisons, if any, will be made, and how this will be done
  - How data from different data collection methods will be integrated (including quantitative data, if any)
  - What you expect as an output (e.g., in-house report, manuscript for peer-review journal, chapter in thesis)

## CHOOSING AN ANALYTIC APPROACH

For analysis of primary data, the choice of analytic approach should be made in the design stages of research. There should be a clear match between the data collection and the analytic approach, with appropriate consideration of the resources that will be available once the data are in hand. Here, we briefly outline three broad approaches to analysis within qualitative research and indicate how they relate to applied thematic analysis.

## Exploratory Analysis

Exploratory analysis is the classic content-driven, inductive approach that most people associate with qualitative research. It can range from a comprehensive ethnographic description of a complex social system to a focused formative assessment intended to inform the design and implementation of a formal evaluation. In exploratory analysis, the emphasis is on what emerges from the interaction between researcher and respondent. The content of that interaction drives the development of codes and the identification of themes. In this regard, the skill and sophistication of those collecting the data determine the extent to which the analysis will be successful. If the data lack richness and nuance, the analysis will similarly be lacking. Thus, in exploratory data analysis it is critically important to conduct quality checks in order to gain some sense of whether what emerges is likely to be more or less inclusive of what may have potentially emerged.

The method most commonly associated with exploratory qualitative analysis is grounded theory. As previously noted, grounded theory was specifically developed for comprehensive analysis of the richness and nuance of exploratory qualitative research. It is the most inductive of the approaches and was designed to guard against interpretive bias in the subjective analysis of textual data. The classic grounded theory approach uses all text generated from a particular data collection event such as an interview. Because of the intensity of the analytic approach, it is difficult to implement with extensive textual data. However, a modified approach can be successfully used in applied thematic analysis for subsets of textual data, if the data collection is appropriately designed.

An example will help to illustrate how applied thematic analysis can be used in a single study to provide both depth and breadth in addressing a research question. This example is drawn from qualitative interview data on the experience of "community" collected as part of a multisite study undertaken in the mid-1990s to identify effective strategies for meeting the social challenges surrounding implementation of HIV vaccine trials. An important foundational question for this research was to understand how community was conceptualized among the diverse U.S. groups with a stake in HIV vaccine trials. Qualitative interviews were conducted with 25 African Americans in Durham, North Carolina; 26 gay and bisexual men in San Francisco, California; 25 injection-drug users in Philadelphia, Pennsylvania; and 42 HIV vaccine researchers across the United States. The interviews covered a range of topics related to the way people experienced and understood community. Using an applied thematic analysis strategy, we extracted all of the responses to a single question: "What does the word community mean to you?" The extracted text was read, and a set of initial codes and definitions developed. Using the iterative codebook development process described in Chapter 3, two coders independently coded all of the text, and intercoder agreement checks were conducted, with adjustments to the codebook and recoding of text as needed. Numeric matrices were then generated to summarize which codes occurred together and then cluster analyzed to identify core elements used to define community, using the strategies described in Chapter 6. The cluster analysis helped identify similarities in the way people defined community as well

as how those definitions varied across the participant groups. A key finding was the identification of a common definition of community shared by all participant groups, as *a group of people with diverse characteristics who are linked by social ties, share common perspectives, and engage in joint action in geographical locations or settings* (MacQueen et al., 2001).

The San Francisco research team simultaneously conducted a more in-depth analysis of community members' descriptions of subgroups of gay and bisexual men. They abstracted substantial sections of the text, including responses to questions on how respondents spent their time, the groups they spent time with or were a part of, the different parts or groups that made up the local gay/bisexual community, and discussions of cultural variation within the community. The analysis identified 32 discrete categories of groups cited by the study participants. They then tested whether these groups met theoretically driven criteria of perceived boundaries, evaluating whether the group was recognized by men who identified with it as well as by those who did not. The researchers identified five major subgroup categories that they labeled leather, men of color, activists, men who go to clubs, and younger men. The descriptive analysis of the responses identified important tensions between the drive to coalesce around common identities and the need to accommodate heterogeneity.

> This understanding offers a different model of community belonging . . . which we will call a multiconstitutive model. In a multiconstitutive community, subcultures and subgroups share their members with a larger community, thereby structuring the gay community as a meta-community formed from the amalgamation of groups within it. These subgroups and subcommunities may not detract from attachment to the gay community, but may be the primary way that many men connect with the diffuse and nebulous thing we call "the" gay community. (Peacock, Eyre, Quinn, & Kegeles, 2001, p. 198)

These findings from analysis at a narrower field of view enriched one of the findings from the cross-site analysis of definitions of community, which was that for gay/bisexual men in San Francisco, a strong sense of shared history and perspective was the most dominant theme, followed by a sense of identity with a specific location, the creation of strong and lasting social ties, established avenues for joint action, and the role of diversity. In contrast to the other participant groups, "Most of the San Francisco participants had thought about community, and many were struggling to reconcile their need for community with a sense of marginalization from society at large" (MacQueen et al., 2001, p. 1935). The more targeted analysis, thus, also added a comparative context to the data, illustrating differences and similarities between groups.

## Explanatory Analysis

Explanatory or conceptual qualitative research uses a combination of deductive and inductive methods and is an increasingly important approach within applied qualitative research. Because such analyses are often used to inform decision making, concerns about the validity of the research are heightened. Such concerns

are in fact justified if the research is premised on a philosophical rejection of objective reality and combined with an interpretation that centers on the researcher's subjective and unchallenged interpretation of the research experience. To address these concerns, we and others have focused on the identification and development of techniques that minimize the potential for what Morse and Mitcham (2002) have described as *conceptual tunnel vision*, that is, "the overcategorization of data, assigning more data to one category than actually belongs, or seeing or justifying most things as being related to, or considered examples of, the concept being investigated" (p. 30). Tunnel vision can also lead to unwarranted exclusion of relevant findings, functioning like blinders that block out data that challenge the framing of particular concepts, thereby making the data functionally invisible. To improve validity in conceptual or explanatory analysis it is critically important to explicitly note when a particular code or theme is linked to the data as well as when it is not linked to the data (see Chapter 3 for more discussion).

Morse and Mitcham (2002) outline a stepwise conceptual research process that includes deconstructing the concept to be explored from the existing literature, developing a skeletal conceptual framework for data collection that focuses inquiry but does not sharply define its limits, and using previous work as a scaffold to explore the internal structure and dynamics of the concept. This is an approach we are currently using in a series of related studies that seek to understand the contribution of concurrent sexual partnerships, or sexual concurrency, to high HIV prevalence rates in southern Africa. Concurrency is generally understood as the situation where a person has two or more sexual relationships that overlap in time. Epidemiologic and behavioral surveys in high HIV prevalence countries in southern Africa suggest that sexual concurrency may be sufficiently prevalent to be an important driver of the epidemic. Computer simulations incorporating empirical evidence from the surveys indicate that the reported rates of concurrency could in fact generate the observed HIV prevalence rates. From this evidence, we developed a conceptual framework of sexual concurrency that served as the *skeleton* (in Morse and Mitcham's terminology) for data collection on both the cultural framing of sexual partnerships and the individual experience of such partnerships in five communities in Lesotho. The results from our Lesotho study, which included 30 focus groups and 93 in-depth interviews, were combined with results from other qualitative research on concurrency in southern Africa; this then provided a more substantial *scaffold* for designing a study of concurrency in seven communities in Zambia. The Zambian study, which centers on 300 in-depth interviews, will ultimately be combined with the data from the 93 Lesotho in-depth interviews and analyzed quantitatively as well as qualitatively. In addition, the Lesotho data are being further used as a scaffold to plan a more comprehensive and largely quantitative sexual network study in that country.

Importantly, the sexual concurrency research in Lesotho and Zambia was initiated by governmental agencies in both countries with support from the Joint United Nations Programme on HIV/AIDS (UNAIDS), the U.S. Agency for International Development, and others. The data were explicitly collected to inform HIV prevention programming and policy. The funding and timelines were constrained, yet maintaining scientific standards was critically important given the way the resulting

data would be used. The data are sufficiently rich to inform theory, but the first priority in our analysis was to explain why and how concurrency occurs, and ultimately provide timely and valid data for informed decision making by national and international policy makers confronting a persistent epidemic.

## Confirmatory Analysis

In the introduction to this book, we outlined the basic differences between exploratory and confirmatory approaches. Here, we discuss considerations for planning a confirmatory analysis.

The first point to recognize is that a confirmatory analysis is testing or assessing a hypothesis; the purpose is to confirm (or reject) a predetermined idea. As such, the procedures involved are more rigorous and less flexible than in a traditional inductive analysis. In fact, there is very little induction involved. In a confirmatory analysis, the conceptual categories are determined *prior* to reviewing the text, and codes are generated from your hypotheses. Creating and defining these codes is an arduous, multistep process. For example, Hirschman (1987) wanted to test hypotheses she had related to gender and resources. In her study, she compared the expression of predetermined themes ("categories") in personal ads placed by men and by women in two large periodicals. The confirmatory nature of the study required going beyond simply creating categories and looking for them in the data. The researcher first verified that all of her 10 categories (i.e., codes) were identifiable in the ads and ensured that items were exhaustive and mutually exclusive. She did this with two other analysts to enhance reliability. Her study also employed a random sample large enough to carry out a comparative statistical analysis to test her hypotheses. In addition, strict procedures were put in place, such as blinding the data analysts to the hypotheses and conducting frequent coding agreement checks, to enhance credibility of her results. Once all of the text had been read, the data coded, and coding agreement assessed, the analysis was finished; there was no room for iteration. The bottom line is that a confirmatory analysis requires more structured procedures at all stages of the analysis process, and in some ways is more similar to quantitative analysis. We refer the reader to other books that provide more detailed instruction for a confirmatory analysis (Krippendorf, 2003; Neuendorf, 2001; Weber, 1990).

## Comparative Analysis: Special Considerations

If you are planning a comparative analysis, the analysis plan should outline how that will be achieved. First, you need to identify the unit of analysis for purposes of comparison. For example, if you are coding transcripts from focus group discussions, there are several options for looking at simple frequencies. You could count the number of transcripts where a code was ever applied, in which case the unit of analysis is the focus group discussion. Alternately, you could count the number of participants in all focus groups whose comments were coded with a particular code, and the unit of analysis is then the individual participant. This approach requires data collection and transcription procedures that allow for the

identification of each speaker. Or, you could count the total number of times a code was applied across all transcripts, in which case the unit of analysis is the text segment. One-on-one interviews require similar decisions. These decisions are described in more detail in Chapters 3 and 6.

The most basic comparative approach in qualitative data analysis is to note themes present in the text from each group being compared and determine which themes are the same and which are different. If the number of data collection events is sufficiently large, you can compare the frequency with which codes are applied to text derived from different populations, subgroups, or categories of people to see if the overall patterning is similar or different. If the sampling procedures for data collection are statistically robust and the number of data collection events sufficiently large, you can evaluate whether differences are statistically significant. You can use graphing or clustering techniques to compare patterns of discourse in the text from different subgroups. The analytic choices for comparing qualitative data are quite rich; Chapter 6 provides an overview of these options.

Finally, consider whether the data collection process is sufficiently similar across the groups to be compared. Here again, an initial scan of the data is important for determining whether the questions asked and the richness of the data collected will permit a meaningful comparison.

## A BRIEF NOTE ON THEORY

As noted, our focus in applied thematic analysis is primarily inductive. Its strength lies in openness to theory building rather than a particular approach. If we consider the explanatory/conceptual, confirmatory, and comparative approaches described above, applied thematic analysis provides basic building blocks. First, as described in detail in Chapter 3, the code-based approach supports the development of taxonomies or classification schemes to aid in sorting and describing the data. A thoughtfully developed codebook serves as a taxonomy, a rich summary description of the range and depth of the data. Second, as described in Chapters 4 through 6, the codebook in combination with descriptive characteristics of the data sources (e.g., participants, observational settings) permits a systematic exploration of relationships in the data as well as comparative analyses of those relationships. Finally, although the analysis process is inductive, it can be incorporated into a multistage, iterative research design where theory in the form of inductive explanation generates hypotheses that can be investigated deductively.

## BOUNDING THE ANALYTIC VIEW

Earlier we referred to *bounding the view* as an important decision in the development of an analysis plan. The decision involves answering three essential questions: (1) What sources of data (e.g., groups, events, individuals) should I include in my

analysis? From those sources, (2) What conceptual/analytic domains should I include? And, (3) What specific questions should I include in my analysis? Remember, when we talk about your analysis we are referring to a specific analysis project that will provide you with the data needed for a given deliverable (e.g., report for funder, journal article, book chapter, conference poster, news brief, etc.). In most research studies, particularly larger ones, there are generally multiple deliverables, and therefore multiple analyses.

While establishing the boundaries of a particular analysis may seem like an important first step, it is in fact contingent on so many factors that it should be one of the last decisions made. Before you can decide what data you will use in an analysis, you need to know what data are available, how the data were collected, how rich the data are, and what gaps exist in the data. That is, you need to know what you are looking at—what is the view provided by the available data?

There is an old joke about a man standing under a streetlight, scanning the ground. A stranger asks him what he is doing and he says he lost his keys and is looking for them. The stranger asks where he lost them, and the man says "Down the street." The stranger then asks why he is not looking down the street and the man replies, "Because this is where the light is." Once you are at the point where you are ready to analyze your qualitative data, you are in the same fix as the man looking for his keys--you can only look at the ground where you let the light of your inquiry shine. This is the hard boundary within which all of your subsequent analysis decisions need to be made.

Bounding the view is essentially a question of what data to use to answer specific research questions. It is driven by the analytic objectives and is a key part of the analysis plan. In its essence, it is a process of data selection by source (e.g., individual, group, observation event), domain of inquiry, and/or question asked. Here we illustrate this process with two examples from our research.

## Prioritizing Analyses From a Large Database

We conducted sociobehavioral research in parallel with a clinical trial that sought to test the safety and potential effectiveness of an antiretroviral for preventing the acquisition of HIV—a strategy generally referred to as pre-exposure prophylaxis, or PrEP. The research sought to facilitate the implementation of the trial in three West African settings and to gather information on the acceptability of PrEP and on the participants' experiences of being in the trial. The research included three phases (pretrial, trial, and post-trial) and lasted more than 3 years. A wide range of participant groups were recruited, including trial participants, clinical trial study staff, policy makers, government and nongovernment service providers, advocates, and media representatives. The research ultimately generated more than 500 in-depth interviews, 28 focus groups, and 78 participant observation events. The data available for analysis also included structured sociobehavioral data from trial participants across 13 time points. Each data collection instrument included multiple domains of inquiry. Adding to the complexity, only one site actually completed the clinical trial; the other two sites closed

prematurely, one due to controversies surrounding the trial and the other due to logistical challenges. Thus, the qualitative data collected during the trial and after trial closure varied greatly from site to site. We were simultaneously confronted with a richness of data and real constraints on our ability to analyze them.

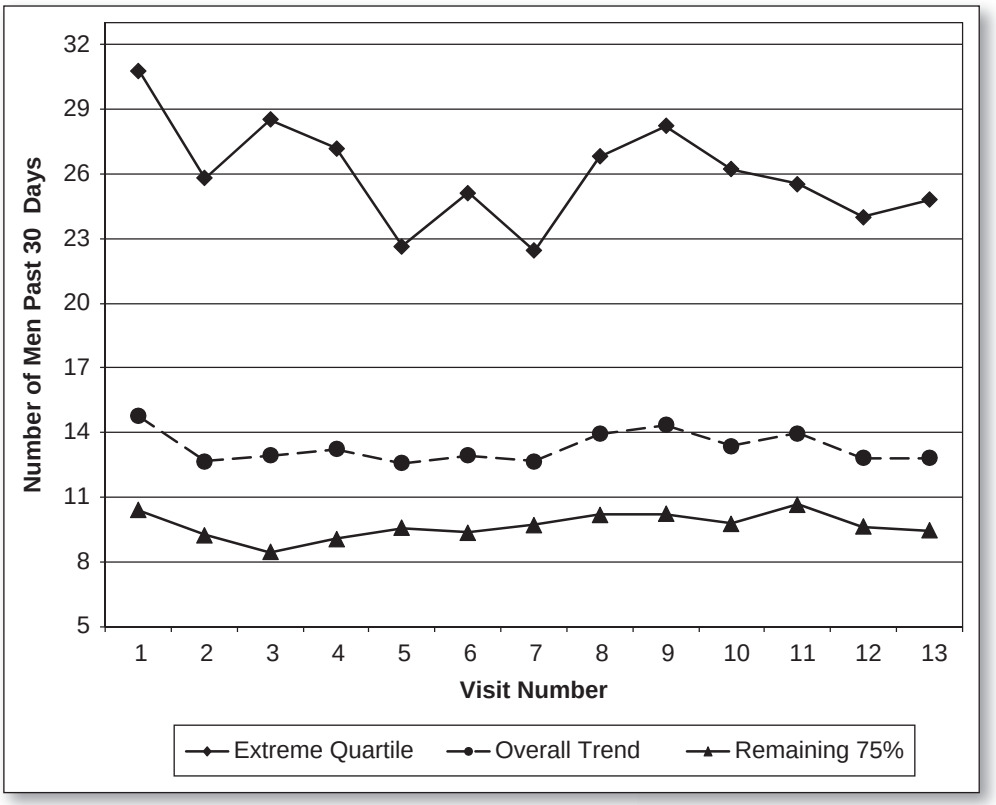
We needed to figure out where to start. First, as part of our on-going quality control, we maintained a comprehensive spreadsheet of data collection instruments and the sites where they were implemented. All transcripts were structurally coded so we were able to determine if there were significant missing data on particular questions. We could also quickly run reports on specific questions to review the richness of the data. We were thus able to identify a short list of key issues of interest and importance to the field and a parallel list of the most robust data on those issues in our database.

We chose three topics for priority analysis. First, we looked at pregnancy decision making among women enrolled in the clinical trial. A number of biomedical HIV prevention trials among women in Africa had encountered difficulties due to high pregnancy rates among participants. Because of concerns about fetal risks, women enrolled in such trials were taken off the study product if they became pregnant. This then decreased the statistical power of the trial and hence the ability of the research to definitively assess safety and effectiveness. The women enrolled in our PrEP trial had high rates of pregnancy, and to understand why, we had added questions about their use of contraception before and during the trial to one of our interview guides. Second, we looked at whether being in the trial was associated with increases or decreases in self-reported risk behavior (a phenomenon known as "risk disinhibition," "risk compensation," or "risk enhancement"). Concerns about risk disinhibition are often raised with regard to biomedical HIV prevention trials, where the actual effectiveness of the intervention is unknown and upwards of half of the participants are receiving a placebo. Third, we looked at the extent to which the trial participants found daily use of the antiretroviral pill to be acceptable. If the women in the trial found it difficult to remember to take the pill, or if they did not like the pill, or if using the pill led to stigma or other problems, there were obvious ramifications for evaluating the effectiveness of the pill for preventing HIV and also for developing viable PrEP programs if the pill were shown to be safe and effective in the clinical trial context. For each of these analysis topics, we strategically chose a combination of qualitative and quantitative data. Below, we elaborate on how we chose data sources to bound the analysis for the risk disinhibition analysis.

The primary question behind the risk disinhibition analysis was: Did the women in the trial change their sexual behavior over the course of the trial? And if so, we posed a secondary question: What accounted for any changes observed? Hence, our objective was to document and explain changes in sexual risk-taking among trial participants. To achieve this objective in an efficient manner, we needed to select only the data most relevant to our goal and underlying research questions. To address the first question, we selected quantitative data collected from all of the study participants at each monthly visit (13 time points, including baseline). We only selected data from the Ghana site because data sets for the other two sites were incomplete because of the early trial closures. Further,

we decided that the quantitative measures most relevant to our analysis were (1) condom use over the past 7 days and (2) the number of different sexual partners in the past 30 days. We graphed data from these two measures and carried out a statistical Growth Curve Analysis to assess change over time and the correlates of change. In Figure 2.1, the line with circle shapes shows the trajectory for the entire study population, with respect to number of sexual partners. For a more detailed account of the analysis and our findings, we refer readers to the original article (Guest et al., 2008).

**Figure 2.1** Number of Male Partners per Month During Trial




In order to address the second question, we looked to the qualitative data and chose for our analysis in-depth interviews administered to trial participants halfway through their participation in the trial. Again, we chose data only from the Ghana site. We also selected just 2 of the 20 or so open-ended questions, from one domain of the in-depth interviews to analyze (note that a targeted analysis like this is made much simpler with the use of a semistructured instrument and structural codes).

These questions, and their subquestions, below, related specifically to our analytic objective and were used to explain trends observed in the quantitative data (we elaborate further on how we integrated the two types of data in Chapter 8).

(1) Has participating in this clinical trial affected your condom use? If so, please explain.

How do you feel about these changes?

What do you think are the reasons for these changes?

(2) How has participating in this clinical trial affected the number of sexual partners you have?

How do you feel about these changes?

What do you think are the reasons for these changes?

The most important point to take away from this example is that we chose only those data that we felt were most relevant to our objectives. We could have included additional data, both qualitative and quantitative, but it would have taken more time and resources to analyze and interpret and would have provided relatively little additional information pertaining to our analytic aims. Always keep in mind the law of diminishing returns. We have often seen researchers unnecessarily bogged down in analytic complexity because they want to include as much data in their analysis as possible. We find it much more effective to break down a large study into several manageable analyses and choose only those data sources and items that will most inform each analysis, as we did in the above example.

## Assessing Strengths and Weaknesses in Data

Other factors to consider when selecting data for an analysis are robustness and validity. Not only should data be relevant to the analytic objectives, they must also be of good quality. A study managed by one of the book's authors illustrates this point well. Tasked with discovering reasons for inaccurate reporting of sexual behavior (a well-documented problem in reproductive health research), Guest and colleagues designed and managed a qualitative study called the Social Desirability Bias (SDB) study (Guest et al., 2005). It was carried out in three African cities—Ibadan, Nigeria; Tema, Ghana; and Gaborone, Botswana. At each site, 30 women at high risk for HIV/AIDS were administered in-depth interviews focusing on how they talk about sexual behavior. The interview guide was composed of 17 questions spread across three conceptual domains—cultural dimensions of sexual discussion, talking about sex in a research context, and perceptions of commonly used techniques in research to enhance accuracy of self-reported behavior (e.g., audio-computer-assisted-self-interviews [ACASI], gender-matching of interviewer/interviewee, etc.). The guide was semistructured, with identical open-ended questions in the same order (but that also permitted inductive probing). The content and format, at least initially, were the same across all three sites. The overall analysis objective was comparative; the plan was to compare data across the three sites, looking for similarities and differences.

As any experienced researcher knows, the real world has a way of humbling research designs. Our SDB study proved to be no exception. The institutional review board (IRB) associated with our U.S. government collaborator for the Botswana site laid down our first bump in the road. After reviewing the protocol and in-depth interview guide, they essentially told us that we needed to ask questions about abstinence and being faithful, to provide a balance for the questions about condom use (the project was reviewed in the political environment of 2003). We had no choice but to oblige if the research were to go forward, even though we felt that adding such questions would disrupt the flow of the interview, potentially offend the participants (many of whom were sex workers), and would provide no useful information. The ethics approval process for the Botswana site also took longer than the other two sites. Because of IRB recommendations and policies, data collection in Botswana ultimately utilized a different instrument than the other two sites and data collection was significantly delayed.

We encountered another challenge when we began analyzing data. Despite pretesting the instrument and monitoring for data quality, data obtained in one of the three domains were extremely weak. The responses were thin and indicated that participants did not really understand the intent behind the questions; their validity was suspect. Faced with these challenges, we had to make some decisions regarding what data to include and how we would parse out the analyses. Our first decision was to analyze the Botswana data separately. We felt that the instrument was too different from the one used at the other two sites to effectively compare the resulting data. We also did not want to delay the analysis of the Nigeria and Ghana data while we waited for data collection in Botswana to be completed. Our second decision was not to include the weak data from the conceptual domain poorly understood by participants based on the likelihood that the views presented there were invalid. We excluded these data from all of the analyses. The result was two separate analyses and subsequent articles—one summarizing the Botswana data (Chillag et al., 2006), the other data from Nigeria and Ghana and including a comparative analysis (Guest et al., 2005) that excluded one of the three domains of questions.

As a side note, we also conducted a methodological analysis with data from the Nigeria and Ghana sites. Because the data from these sites were robust and could be meaningfully aggregated, we did our substantive analysis in a stepwise manner that documented codebook development and application of codes after the analysis of every 6 in-depth interviews, until all 60 interviews were coded. Based on this analysis and audit trail, we identified points of data saturation within and across the data sets from the two sites. The resulting article provides one of the few evidence bases for estimating nonprobablistic sample sizes (Chillag, Guest, Bunce, & Johnson, 2006).

In sum, you need to have an analysis plan, but be prepared to be flexible and responsive to exigencies that present themselves on the ground. Data may not always be of good enough quality to include in your analysis (which is *not* to say that you should not include negative cases or present contradictory evidence in your findings. We argue against such exclusions in Chapter 4). At the same time,

make the most of your data. A data set may address more than your primary analytic objectives, and responding to these secondary or tertiary objectives may turn out to have more impact than one might imagine. The secondary, methodological, objective of identifying points of thematic saturation in the example above proved to be an important contribution to the field of qualitative research methods, despite the challenges encountered during analysis of the primary objective.

## SUMMING UP

The key to developing an effective analysis plan to answer a research question is to identify the discrete activities that need to take place in order to achieve your desired outcome. The Appendix summarizes the four key elements that are generated in the analysis of qualitative data and thus provides an overview of the process. Each element corresponds to a fundamental type of information collected in the research process. First, there are the characteristics of the sources where answers to research questions are sought. As noted previously, the sources may be individual people (including the research team) or groups of people or institutions. Second, there is the primary information collected from the sources, for example, the transcripts of interviews or focus groups. Third, there is information generated to assist in the interpretation of the primary information, referred to here as secondary information to distinguish it from what is obtained directly from the sources. This secondary information includes codes, code definitions, and coding notes; the linkages between the source information and the codes, definitions, and notes; and descriptive summaries of the source information such as word frequencies, code frequencies, code co-occurrence matrices, and the like. Finally, there is information about the characteristics of the coders who generate the secondary information. This framework emphasizes an object-oriented definition of data where data are defined as any digital representation, including photographs, graphic displays, video, sound, text, and numeric data. All of these representations are essentially database objects or artifacts created by people. Some are created to describe or explain other objects in the database (the who, what, where, why, and how of the objects). By considering the relationships among and the content within each of these elements of a research project, qualitative researchers can systematically organize their data to make analysis and the reporting of results more efficient and reliable (MacQueen & Milstein, 1999).

Although such an overview is helpful, it is important to remember that the process is dynamic and iterative. The goal of the analysis journey is not reached by taking the shortest path between research question and answer. It is an exploration of little-known territory, and at the end of exploration, you should have a map that others can use to more efficiently understand and explore the same territory.

## REFERENCES

Chillag, K., Guest, G., Bunce, A., Johnson, L., Kilmarx, P., & Smith, D. (2006). Talking about sex in Botswana: Social desirability bias and its possible implications for HIV prevention research. *African Journal of AIDS Research, 5*(2), 123–131.

Chillag, K., Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods*, *18*, 59–82.

Guest, G., & MacQueen, K. (Eds.). (2008). *Handbook for team-based qualitative research*. Lanham, MD: AltaMira Press.

Guest, G., Bunce, A., Johnson, L., Akumatey, B., & Adeokun, L. (2005). Fear, hope, and social desirability bias among women at high risk for HIV in West Africa. *Journal of Family Planning and Reproductive Health Care*, *31*(4), 285–287.

Guest, G., Shattuck, D., Johnson, L., Akumatey, B., Clarke, E., Chen, P., & MacQueen, K. (2008). Changes in sexual risk behavior among participants in a PrEP HIV prevention trial. *Sexually Transmitted Diseases, 35*(12), 1002–1008.

Hirschman, E. C. (1987). People as products: Analysis of a complex marketing exchange. *Journal of Marketing, 51,* 98–108.

Krippendorf, K. (2003). *Content analysis: An introduction to its methodology.* Thousand Oaks, CA: Sage.

MacQueen, K. M., McLellan, E., Metzger, D., Kegeles, S., Strauss, R., Scotti, R. Blanchard, L., & Trotter, R. (2001). What is community? An evidence-based definition for participatory public health. *American Journal of Public Health*, *91*(12), 1929–1938.

MacQueen K. M., & Milstein, B. (1999). A systems approach to qualitative data analysis and management. *Field Methods*, *11*(1), 27–39.

Miles, M., & Huberman, M. (1994). *Qualitative data analysis* (2nd ed). Thousand Oaks, CA: Sage.

Morse, J., & Mitcham, C. (2002). Exploring qualitatively derived concepts: Inductive-deductive methods. *International Journal of Qualitative Methods*, *I*(4), Article 3. Retrieved from http://ejournals.library.ualberta.ca/index.php/IJQM/article/view/4589

Neuendorf, K. (2001). *The content analysis guidebook.* Thousand Oaks, CA: Sage.

Peacock B, Eyre, S., Quinn, S., & Kegeles, S. (2001). Delineating differences: Sub-communities in the San Francisco gay community. *Culture, Health & Sexuality*, *3*(2), 183–201.

Schensul, S. L., Schensul, J. J., & LeCompte, M. D. (1999). *Essential ethnographic methods: Observations, interviews, and questionnaires, ethnographer's toolkit* (Vol. 2). Lanham, MD*:* AltaMira Press.

Weber, R. (1990). *Basic content analysis: Quantitative applications in the social sciences.* Thousand Oaks, CA: Sage.

## EXERCISES

Using the list of items to consider in an analysis plan provided in Table 2.4, write a comprehensive analysis plan for an existing or anticipated thematic analysis. Be as specific as possible. Draw a flow diagram outlining the procedures and actors involved. Review your plan and check to see if you (or we) have missed anything that might be important to note.

## ADDITIONAL READING

Dey, I. (1993). Finding a focus. Q*ualitative data analysis: A user-friendly guide for social scientists* (pp. 63–73). New York: Routledge.

Miles, M., & Huberman, M. (1994). Focusing and bounding the collection of data. *Qualitative data analysis* (2nd ed., pp. 40–49). Thousand Oaks, CA: Sage.

Patton, M. (2000). Qualitative analysis and interpretation. *Qualitative research and evaluation methods* (3rd ed., pp. 431–534). Thousand Oaks, CA: Sage.