CDC Excel Tool for Thematic Analysis- User guide and tutorial

Contents

This guide was developed by the **CDC NCEZID Social and Behavioral Science Team** to accompany its **Excel Tool for Thematic Analysis.** To request a copy of the Excel Tool, or for more information, send requests to textanalysistool@cdc.gov.

Introduction

This guide will help you use the **CDC Excel tool for thematic analysis** (we will call it "the tool"). The tool is an MS Excel file that that has been programmed so you can: 1) plan your inquiry and set up the analysis; 2) code important segments of text according to their meaning; and 3) identify important themes emerging from the coded data ("thematic analysis").

While you can use the tool to organize and analyze any kind of text, the tool was originally designed for use in an epidemic response. There is a coding scheme that was developed and refined after coding text for numerous different epidemic responses, providing codes common to most epidemics, and making it possible to customize the codes as needed. A second coding scheme is provided that is more appropriate for other types of health emergencies, such as natural disasters, where health and safety is still a concern, but where there may be several different kinds of threats present at the same time. Lastly, a third, completely blank coding scheme structure is provided for those who want to create their own coding scheme.

If you are involved in an epidemic response or other type of health emergency and you have notes or transcripts from meetings or conversations that you want to organize and analyze, this tool can help you. You can also use this tool to identify themes from social media posts and comments. With this tool, the analysis you will do is *qualitative*, which means that you are trying to identify the main ideas expressed in the text, and to describe them as fully as possible, without trying to determine the percentage of the population that holds each viewpoint or attitude. It may be helpful to think of this type of analysis as answering the question, "What are we hearing from the public versus what are we NOT hearing?" rather than, "What percentage of the population holds these views?" Qualitative analysis of this type is particularly valuable because it often reveals perspectives and views that we have not heard before or have not thought to ask about in surveys or interviews. This kind of analysis is often the first step in developing a survey or more quantitative study.

You do not have to be experienced in text coding or thematic analysis to use this tool, but you should have at least a basic knowledge of MS Excel.

Chapters 1 and 2 describe how the tool is set up, with three phases (set-up, coding and creating themes), and 3 to 4 worksheets used in each phase. Chapters 3 – 5 take you step by step through how to use the tool. Chapter 6 covers a few additional topics for advanced users. You may want to read chapters 1 and 2 and watch the instructional video tutorials first to familiarize yourself with the tool, and then read through the step-by-step instructions in chapter 3 through 5 after that.

Video tutorials – We have created six brief video tutorials, one to provide an overview of the entire tool, and five more to walk you through the process. When you see a video tutorial button, you can select it to view a brief video tutorial. Follow this link to access a playlist of all six of the video tutorials.

Please note that in MS Word, you will need to hold down the CTRL key as you select the green button for the video tutorial to activate. While you can learn all of the steps just by viewing the video tutorials, this guide is a more detailed version of the information contained in the video tutorials.

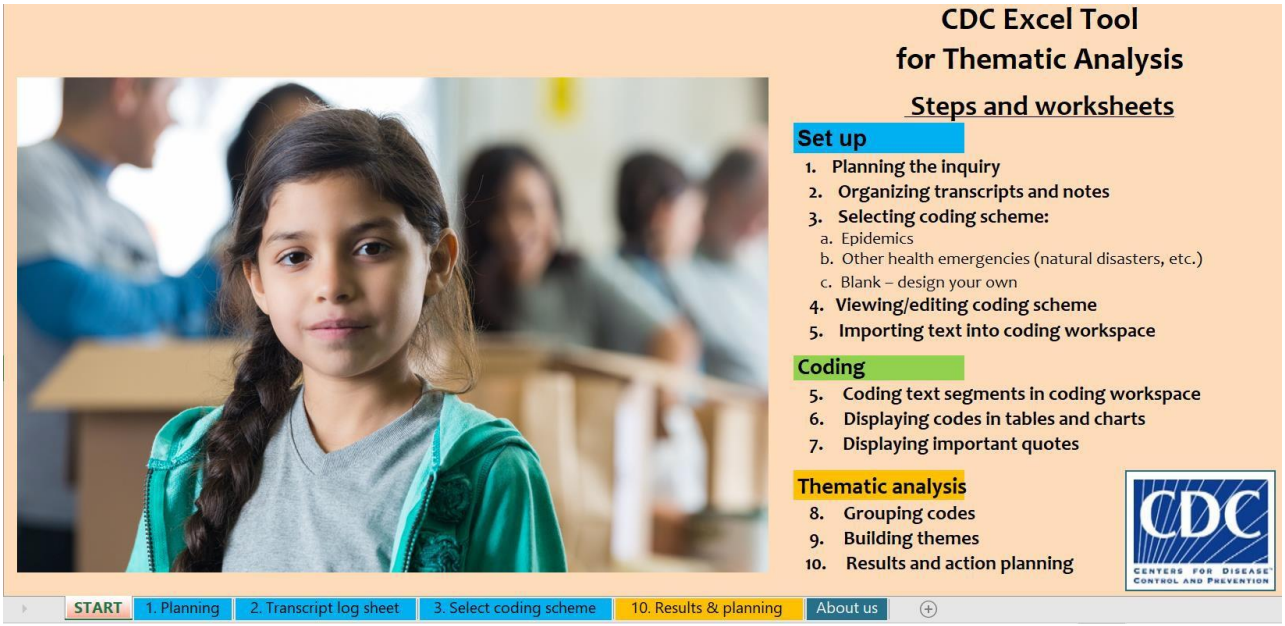
If you would like to view the introduction video tutorial now, you can access it here (while holding down CTRL button): video tutorial 1. If you are reading this guide in print, you can search the internet for "CDC Tool for Thematic Analysis YouTube".

1. Overview: Tool format and structure

1.1. Format

The tool is organized around three phases: I) setting up your text and preparing for coding; II) coding your text, and III) building themes from the codes. Figure 1 below shows the START page of the tool. You can see how the 10 worksheets are used in sequence in the three phases. Each phase is highlighted in a different color (I: set up/blue, II: coding/green, III. Thematic analysis/yellow).

Figure 1.a. START page of Excel tool, showing phases and worksheets



Phase I: Set up - Formatting your notes, transcripts, or social media comments in a text editing program like MS Word, then pasting it all into MS Excel. You will do this in four steps:

1. Write out in your own words what you want to learn from the analysis
2. Select one of two pre-existing coding schemes to use with your text, or select the blank scheme and enter your own coding scheme
3. Format your text (notes, transcripts, and social media posts) in a Word document
4. Paste the text into the Excel coding workspace

Phase II: Coding - Using pre-formatted drop-down menus to code sections of your text, and then when you are done, view all the coded text sections in groups by code in pivot tables. "Coding sections of text" simply means assigning a label (we call these labels "codes") to briefly describe that section of your text. You will perform five steps to complete this phase:

1. Review each sentence in your text and decide if you want to copy the entire sentence to the next column for coding. Typically, we code one sentence at a time. You can also duplicate a sentence and code it two different ways. If you decide to not code all of the text, then you will just copy over the part you want to code.
2. Review the text and decide whether it represents one idea or more than one. If more than one, consider duplicating the sentence so that you can code it two or more different ways. Double

code when needed (this means to duplicate the segment row, then code each row differently as needed)

3. Use the pre-formatted drop-down menus to assign codes to text segments
4. Open pre-formatted pivot tables, refresh data and view results

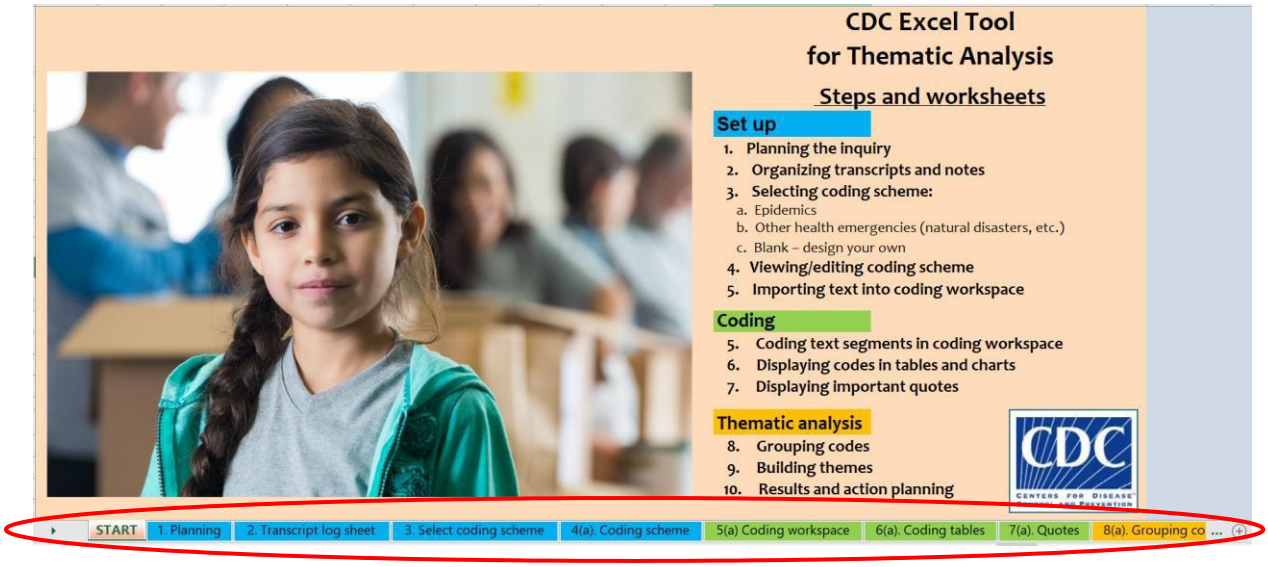
Phase III: **Thematic analysis** - Using the text sections grouped by the pivot table, create themes that expand on codes or groups of codes. Identify specific quotes or text segments that best represent each theme. You will perform five steps to complete this phase:

1. Select a pivot table that shows your most important coding scheme(s)and paste into the thematic analysis workspace
2. Use the workspace to group similar codes together
3. Write a sentence next to each grouping that says what each sentence means, in your own words. This is a way of simplifying what was said.
4. Identify key quotes to illustrate your thematic statements
5. Interpret and decide what the themes mean for your work. Think about it compared to your original goals on the **planning (1)** worksheet. Is it related to your question? Does it suggest an answer to your question?

1.2. Structure

The tool contains several worksheets numbered in the order that you will use them. The worksheets are also color coded according to the phase in which you use them. The **coding workspace (5)** worksheet is used in both the set up phase and the coding phase. It is colored blue with the set up phase since that is when it is first used. As shown in figure 1 below, step 3 in the set up phase is to select a coding scheme from among three, which are labeled a, b and c. Once you make this choice, the worksheets 5 through 9 will appear, showing either a, b or c in the worksheet name.

Figure 1.b. START page of Excel tool, showing worksheet tabs



The following briefly describes the purpose of each worksheet:

- **START** – Open this worksheet when you begin. On this page (shown above), you will see a brief outline of how the tool is structured, with three phases (set up, coding and thematic analysis), and numbered actions under each phase. The number next to the action also corresponds to the number of the worksheet that you will use with that action.
- **1. Planning -** This worksheet helps you think through why you are collecting and analyzing community listening information. Have you detected possible needs, concerns or issues in the community that you want to understand better? How will having this information help you meet community needs better? Spending some time at the beginning of the process thinking through how the results will be used will be tremendously helpful in assuring that the time spent (both yours and that of community participants), results in benefit to the community.
- **2. Transcript log sheet –** This worksheet is a place to keep track of your notes and transcripts by listing when the listening occurred, whether it was a focus group or some other type of listening, how many people participated, and including a brief note about the kinds of people who participated.
- **3. Select coding scheme –** On this worksheet you will select one of three coding schemes for coding your text. When you make a choice (by selecting the "epidemic coding scheme," "other health emergency coding scheme," or "blank coding scheme" button), this will populate the next worksheet (coding workspace) with appropriate coding options. If you select the "blank coding scheme" you will be able to enter your own codes. The other two coding schemes can also be edited.
- **4. Coding scheme –** Once you select a coding scheme, this worksheet will automatically open to show all of the codes in the coding scheme you chose. You can add and edit codes in any coding scheme. All of the coding schemes are set up with two levels, meaning that first you will assign a broad code (or category) to the text segment, and then you will assign a second, more specific code. This makes it easier to work with the data later. For this worksheet and worksheets 5 through 9, the letter a, b or c will appear on the worksheet tab to indicate which coding scheme it goes with.
- **5. Coding workspace (a) (b) (c) –** This worksheet has a space for your text to be imported (in the set-up phase) and drop-down menus for coding your text (in the coding phase).
- **6. Coding tables (a) (b) (c) –** Once you have finished coding your text, you can open this worksheet, refresh the data, and it will repopulate the tables and charts displaying your codes in a number of different ways.
- **7. Quotes (a) (b) (c) –** After you have coded your text, it can also be very helpful to look at your text segments grouped underneath the code. It can help you to review your coding, and also help identify characteristic quotes that you might want to share in your final report. This worksheet is programmed to show these coded segments with the associated code. Buttons let you select which codes to view.
- **8. Grouping codes (a) (b) (c) –** Now that you have reviewed all of your coded text, the final phase is the thematic analysis. This worksheet is for the first step in that analysis, in which you will group different codes that relate to a single idea together in boxes. This will help you start to see patterns and themes.
- **9. Building themes (a) (b) (c) –** While the **Grouping codes (8)** worksheet has you group codes in boxes set side by side, worksheet 9 has you take the contents of those boxes and paste them into

the first column, one below the next. Once ordered vertically, you proceed across the table, left to right, describing the meaning of each code in more detail, summarizing groups of codes and finally, synthesizing what you have written into a paragraph. This sheet reminds you of your original questions so that your synthesis relates to your original aims.

- **10. Results and action planning –** Worksheet 10 brings you back to the questions from worksheet 1, how will you use the answers to your questions to take action? Further, who are the actors that help you take action?
- **About us –** This final worksheet names the CDC team and the individuals responsible for the creation of the tool, and provides an email address for more information.

2. Examples of different kinds of listening texts

On many of the worksheets you will see examples of possible entries in italics in the first few rows. In addition, for the coding and analysis steps, we will use example texts that come from imagined listening sessions about a made-up disease, "disease x." These three examples show what different types of texts you might work with look like. A brief description of each type of text follows, along with the example text itself.

- **Example 1.** A transcript of a **focus group** with members of the community. A transcript is a document that comes from an audio recording. This kind of text is going to be very complete. Everything that was said will be in the transcript, even if it is off topic. With a focus group, the facilitator asked questions designed to stimulate free flowing discussion to reveal as much descriptive detail as possible. You will have statements, questions, and comments from several different people. Some focus group transcripts identify the different speakers (e.g., "participant #1, #2, etc.,") while others will leave the text without any identification. More advanced users may want to distinguish speakers within a group but in our training, we will not make that distinction.

  Example **focus group** text:

  > "Facilitator: What have you heard about *disease x*? Respondent 1: Kid will have delays as they get older -- it causes problems for them. And I worry about long term effects for adults too because no one seems to know about that. Respondent 2: I learned yesterday that people can get reinfected even after having it one time Respondent 3: I didn't know that. You know that kind of thing that makes me really scared. I thought it was like the chicken pox.
  >
  > Facilitator: Do you have any concerns or fears? Respondent 1: I mean, at first, I did but now that I'm vaccinated, and my kids are vaccinated? And we just been home I mean, ninety percent of the time home alone only the kids go to school, and we get groceries delivered so we don't a lot of risk.
  >
  > Respondent 3: I'm less concerned about getting infected but about being treated for it. Facilitator: What specifically concerns you about treatment? Respondent 3: Well, I know it's very hard to get a doctor's appointment and hospitals are incredibly busy so if we got sick where would we go? And how do you know that you're sick enough to need treatment? There isn't much information available about treatments for milder illness or how to recognize that mild symptoms are becoming severe."
  >
  > Facilitator: And in regard to *disease x*, what you heard about like where did you gather all your information from? Respondent 1: Every time I -- by the clinic. I pick up the pamphlet and that stuff."

- **Example 2.** Notes from a **semi-structured interview**. In this kind of interview, the interviewer asks questions, but the respondent answers them without any prescribed response choices. They answer in their own words. This kind of interview also seeks to draw out descriptive detail, but from one person. In this case, the text is notes taken by an observer, so it will be less detailed than a transcript. It still contains some quotes.

  Example **notes from a key informant interview** with a community leader**:**

  "What people are doing to protect themselves:

  - Hand washing
  - Hard to get hand sanitizer
  - Thinks HS is better than soap

    Added: should have video tutorials about hand washing, multiple languages.

  Best ways to share information with community:

  - "Best is to have a very authentic conversation."
  - Need one platform where everyone can go and get ALL the information
  - Lack of internet problem for some, but bigger problem is not speaking English. Need multiple languages.

  How disease has impacted the community:

  - "It's hurt everything and everyone. People can't get to their usual doctors because so many of the local doctors have been pulled away to work in hospitals."
  - "So many people staying home so restaurants and businesses are empty."
  - People don't go near each other on sidewalk, fear
  - "I know that we needed these measures to fight the disease but I worry that all the other stuff we're not paying attention to is going to cause problems in the future.""

- **Example 3. Comments on a CDC Facebook post.** In this case, the comments were taken from a social media site. These comments are much briefer, and there would be no way to tell if one person had made more than one statement.

  Example **posted comments on a CDC Facebook post:**

  "CDC: Vaccines have been saving lives and preventing illness and disability from devastating diseases for over a century. Today, *disease x* vaccines are helping protect people against severe illness, hospitalization, and death. Vaccines are free and available at your local pharmacy, talk to your healthcare provider about getting yourself and your family vaccinated today!"

  Comments:

  - The vaccine accelerates certain diseases in the body, that's why we are afraid
  - *Disease x* will not be eradicated because people who are cured continue to infect others because people are selfish and won't stay home when they are sick
  - *Disease x* does not exist; it's a way to control people under unfair government regulations
  - I do not believe that *disease x* exists because I don't know anyone who's ever been infected to believe in the disease
  - What is the point of receiving this vaccine when even people who are vaccinated still die?
  - When I tried to get my son vaccinated first there were no appointments available and when we finally got an appointment they tried to charge me for it!!"

3. Phase I. Set up: preparing for coding and analysis

Click on this link to see a short video tutorial that reviews the set-up phase, including planning the collection, tracking the types of information being collected and selecting a coding scheme, using worksheets 1 – 3. video tutorial 2.

1. **Planning the inquiry** – Open the **Planning (1)** worksheet and think about why you are doing community listening. What do you hope to learn? Use the questions column to write down your thoughts. These questions will come from your initial understanding of the situation, maybe even from clues that you have that community participation is not as good as it could be. For example, in past epidemics, social scientists identified the following "clues" that public participation and support for infection control efforts were experiencing problems:

- **Zika, Puerto Rico** – Initial recommendations that pregnant women wear long sleeves and long pants and stay indoors to avoid Zika infection through mosquito bites was not being widely adopted. Community listening revealed that long sleeves and pants were extremely uncomfortable in the climate, and that many homes did not have screens or air conditioning. Further, women felt that the recommendations placed all of the burden on pregnant women alone, leaving them unsupported. This information led to strategies for families and communities to participate in removing mosquito breeding sites from neighborhoods, and to emphasize that protecting unborn children was everyone's responsibility. [C Prue, 2016, unpublished]
- **Ebola, Democratic Republic of Congo –** Early in the 2018-2020 outbreak there were reports of families burying their loved ones who had died of Ebola secretly in order to avoid the mandated safe burial practices by local government or the Red Cross. Community listening revealed that "safe and dignified burial" practices did not allow for religious and family leaders to perform important burial rites and blessings, and further that keeping the body "hidden" in a sealed body bag raised suspicions that the body had been stolen and replaced with another or that certain body parts had been removed. With this information, burial teams were able to make their burial process more flexible, allowing families a say in how the burial would be performed. In addition, the Red Cross began using transparent body bags, which allowed families to verify that their loved one's body had been treated respectfully. [Earle-Richardson et al., 2021]

- **COVID-19, U.S. meat packing industry –** Early in the COVID-19, a high rate of infection was detected in meat processing facilities. Community listening revealed that many workers were very anxious to return to work but were unclear about new infection control practices and leave policies. Part of the reason was that workers spoke a wide variety of foreign languages, making updating safety precautions and company health and safety policies challenging. It also uncovered a network of support agencies and online platforms that could assist in rapidly disseminating new information in a range of languages. [C Prue, G Earle-Richardson, 2020, unpublished. In all three of these examples, public health officials saw initial signs that response efforts weren't proceeding as hoped, and they needed more information about why. Apply this thinking to your situation. What initial signs are you seeing or hearing that make you want to investigate further? How

could this knowledge solve a problem or make infection control (or other health promotion strategies) work better?

Keep in mind that when you analyze text of this kind, you are trying to get a picture of people's experiences, concerns, and viewpoints as they describe them. You are NOT going to be able to find out what percentage of people feel one way or another or associate the comments with different groups in a precise way. Surveys are more appropriate for that kind of information. It is important to keep these limitations in mind when you write down your questions so that you don't pose a question that can't be answered with thematic analysis. The table below provides some further examples of questions that might be answered with thematic analysis, and which would be better answered with a survey or other quantitative methods.

Table 1. Comparison of analytic questions that can be answered with qualitative and quantitative analysis

Area	**Questions answered through qualitative analysis** (e.g., thematic analysis)	**Questions answered by quantitative analysis** (e.g., survey)
Disease testing	What kinds of concerns are mentioned around testing for [outbreak disease]?	What percentage of respondents feel that testing is too expensive?
Trust in health authorities	Is mistrust of health officials a theme that comes through when talking about the government-led response?	What age groups are least likely to trust health authorities? Has this level of trust changed over time?
Root causes of negative perception of government- led response	When respondents describe negative feelings toward the government-led response, do they talk about negative personal experiences, or things they have heard in the media, or both?	Among respondents who hold a negative view of the government-led response, what it is most commonly cited reason? Is this the same among males and females?

Click here to learn more about thematic analysis and qualitative analysis

After typing in as many questions as you have, use the next column (Uses) to ask how the answers to each question will contribute to the emergency response. Next, moving to the "Implementors" column, think about who in the response is best positioned to effectively implement actions recommended by the information you uncover? In the last column (Demand for information), consider whether there is currently interest from these people in your inquiry or in this kind of information. Have you communicated with them? If so, have you determined that they are interested in this information, and that they don't already have the answers?

This planning step is critically important for assuring that the time you spend (and the time spent by the community) produces useful information that contributes to ending the crisis.

2. **Organizing transcripts and notes -** While designing your strategy for listening (e.g., focus groups, town hall meetings, listening during existing gatherings, etc.), is a large topic, and is beyond the scope

of this analysis tool guide, the **Transcript log sheet (2)** worksheet can help you keep track of the characteristics of the individual or group that is represented by each listening text as well as which listening texts have been formatted and coded. For example, you may want to have a record of when each listening text was collected, the number of participants, the age range, sex, or some other characteristic important to you. This is particularly useful when you are analyzing several different listening texts at the same time.

Here is a brief description of some of the pre-programmed columns:

File ID - The first column will automatically a file ID (starting with #1). You can also enter your own ID number if you prefer. This ID number appears in the Coding workspace (5) worksheet next to all of the coded text segments from that listening text.

Collection Date - This column is where you can record the date on which the information was collected. When you fill out this column the "Total Number of the Transcripts" row on the top will update.

Type of event - Here you select from a drop-down list to assign a "type of event." You choose from a long list of listening formats, such as focus groups, social media posts, complaint box, etc.

Transcript / Notes - In this column you can choose whether the information is a full transcript of the event or if the text are notes from the event. This can be useful when you want to have an indication as to whether you captured all of the comments in people's exact words, or whether you took notes, which typically are shorter and not necessarily exactly what participants said. With notes you will be more cautious about using quotes in your final report, since they may not be exactly what the person said.

Number of Participants - This column asks how many people participated in the listening session or event. When you enter the number, the "Total Number of Participants" row above will update automatically.

File Name - This column is for entering the name of the listening text file. This is helpful because you can match the details displayed here to the text file that is stored outside of the Excel tool.

Formatted for Importation? - This yes/no box allows you to note whether you have completed formatting the text file so that it is ready to be pasted into the Excel tool.

Coding complete? - This yes/no box allows you to note whether you have completed coding the text. This allows you to remember your work progress if you are formatting and coding multiple listening texts. As you change the values in the formatted and coded columns, the percent formatted and percent coded cells above will also update.

Participant Type – This last column is not formatted; you can enter any type of information that is meaningful to you about the types of participants. This is to help you remember who provided the listening information.

You may also want to add your own columns for additional information about your listening texts. To do that, right click on the column next to where you would like to add a new column, go to "Insert" and then choose "Table Columns to the Left." And provide a new name for your column.

3. Selecting coding scheme – Next, you will open the **Coding scheme (4)** worksheet, and that will open to three boxes, each with a blue button with it. The three buttons are labelled, "Epidemic Coding scheme," "Other Health Emergency Coding Scheme" and "Blank Coding Scheme."

Understanding coding and how to choose a coding scheme - Before going ahead with selecting a coding scheme, it will be helpful to spend a few minutes talking about what coding is, and what are the important considerations in choosing a coding scheme.

"Coding" is a process of reading each sentence and then selecting a label or "code" that best describes what the sentence is about. Once you have done this with all of the important sentences in your worksheet, the tool will display your codes according to the frequency of each being assigned. This will help you understand the text better, and you will use it to build your themes.

There are many different ways to code text. For example, you can assign codes that signify the meaning of the sentence (descriptive coding), or you could assign codes that indicate the type of emotion revealed by the statement or question (sentiment coding), or you could even code sentences according to some specific characteristic that is important to you (e.g., perceptions of unequal treatment, requests for services, etc.). There are many different possibilities, depending on the goals of your analysis. In this tool, we are mainly focusing on the simplest and most straightforward coding—descriptive coding. Experience has shown that a two-level coding system in which the first, "broad code" indicates the general topic being spoken about, and a second, more specific code that provides more detail is very helpful.

In this tool we have created a two-tiered descriptive coding scheme that you can use to code text related to an epidemic response. It has been developed through years of coding during epidemic responses. Both levels of coding have been used for coding during several epidemics and have been found to be consistently applicable. More recently, we have developed a second version of the coding scheme that is oriented toward health emergencies in which there may be multiple health threats in addition to infectious disease transmission. This coding set is a little broader and has not been tested in actual health emergencies. Both coding schemes can be customized to your needs. You may review the codes available under each of these schemes in the Appendix of this guide.

Because many times analysts prefer to create their own coding schemes based on their particular interest area and their text, we have also created a third coding scheme which is entirely blank, so that you can code using a scheme that you have created entirely on your own if you choose.

Select "Epidemic Coding Scheme" - In order to familiarize yourself with the tool and see some examples, select "Epidemic Coding Scheme" and you will be automatically redirected to the **Coding scheme (4)** worksheet. This worksheet displays all the codes that will be available when you code using this coding scheme. This coding scheme has been developed over years of coding text from communities during a range of different epidemic responses.

Broad topic codes: The first column shows the "Broad Topic Codes," meaning the first set of codes that you will use to code the broad area in which your text segment fits. From column c and then continuing

to the right, you will see each broad topic code again with a list of specific codes below it. The broad codes show the five main categories in which comments tend to fall, with a six "other community issues" area for other issues that may appear.

Specific codes: These specific codes will appear in a drop-down menu on the coding workspace worksheet after you select a broad code.

Optional "Action Flags" secondary coding scheme - In column I, at the far right, you will see a column named "Action Flags." This is an additional coding scheme that you can use at the same time as the two-level descriptive coding scheme. This coding scheme alerts you that a particular text segment needs immediate follow up. You will have to decide within your own specific situation what kind of comment requires immediate follow-up, but experience has shown that reports of insecurity or abuse in the response typically would need to be reported, as would threats of violence. Other, less dangerous "flags," but still possibly requiring immediate action could include comments indicating that misinformation is being disseminated, questions that the response would want to answer quickly and publicly, other indicators of an operational problem in the response that leaders might not be aware of, or other suggestions for action.

All the cells on this sheet can be edited simply by typing over the text in the cell, so you can customize the action flags list to fit your situation. You can also edit any of the specific codes by typing over that cell. You can also add specific codes by right clicking on a cell in the middle of the list (*not* at the end of the list), selecting "insert cell," and then "shift cells down." We don't recommend adding to or changing the broad codes, since the code names are part of the programmed link between the broad codes and the specific codes. If your desired coding scheme differs substantially from the first two coding schemes, you can select the third, blank coding scheme and make your own. See section 6. Selected Advanced User Topics for information on how to modify the coding schemes.

Next, select "Hide Coding Scheme" to go back to the **Select coding scheme (3)** worksheet. Try selecting the "Other Health Emergency Coding Scheme" button and review this coding scheme. You will see that the coding scheme is set up the same way as the epidemic coding scheme, only with slightly different broad topic codes that reflect an emergency with a range of different possible hazards, including some not related to infectious disease. This scheme has been reviewed by humanitarian health emergency specialists but has not been used yet in actual emergencies. As with all of the coding schemes, you can edit the specific codes and the flags, and you can specify additional codes through the use of "other community issues" broad code and adding relevant specific codes to that list.

4. Importing text into coding workspace – The final step in the set-up phase is bringing your text into the coding workspace. Click on this link to see a short video tutorial that will show you how to format and import your data into the tool. video tutorial 3.

You will begin with your text in a word processing file, such as Word. You may want to just import one text to code it, or you may choose to import several texts and code them and analyze them for themes all together. To bring the text into the **Coding workspace (5)** worksheet, you are going to simply paste your text into the "Text Sentences" column (column F). But before you do that, you will need to format your text so that each new sentence (or line of text in the case of notes) begins a new paragraph. This is important because Excel reads the new paragraph symbol as the beginning of a new row. At this point in the process, you will assign each sentence to its own row. Once you have pasted the text into the

tool, you will have the chance to break the sentences into smaller segments (such as phrases), or to group more than one sentence in a single cell for coding if you choose.

When you first paste your text into MS Word, the document may not have any formatting, or may be formatted into paragraphs. Here is our first example, the transcript, before we format it:

Figure 2. Example of text before formatting

Facilitator: What have you heard about *[disease x]?*[2] Respondent 1: Kid will have delays as they get older -- it causes problems for them. And I worry about long term effects for adults too because no one seems to know about that. Respondent 2: I learned yesterday that people can get reinfected even after having it one time Respondent 3: I didn't know that. You know that kind of thing that makes me really scared. I thought it was like the chicken pox. Facilitator: Do you have any concerns or fears? Respondent 1: I mean, at first, I did but now that I'm vaccinated, and my kids are vaccinated? And we just been home I mean, ninety percent of the time home alone only the kids go to school, and we get groceries delivered so we don't have a lot of risk. Respondent 3: I'm less concerned about getting infected but about being treated for it. Facilitator: What specifically concerns you about treatment? Respondent 3: Well, I know it's very hard to get a doctor's appointment and hospitals are incredibly busy so if we got sick where would we go? And how do you know that you're sick enough to need treatment? There isn't much information available about treatments for milder illness or how to recognize that mild symptoms are becoming severe. Facilitator: And in regard to *disease X*, what have you heard about things like where did you gather all your information from? Respondent 1: Every time I -- by the clinic. I pick up the pamphlet and that stuff.

There are two options for formatting your text at this stage. If you have a small amount of text data (1-2 pages total) we recommend you use **Option #1:** review the entire document and manually insert a carriage return ("enter") at the end of each sentence.

Option #2 is a bit more involved, but if you are working with transcripts longer than 2 pages, it can save you a lot of time. You will use the "find and replace" function in Word to replace all the punctuation marks in the text with new paragraph symbols ("enter"). (Option 2 assumes that the end of each sentence in your text is marked by a period, question mark or exclamation point.)

Steps for automatically adding a paragraph mark to the end of each sentence:

1. Identify common terms or words in your text transcript that have periods and replace these first. For example, if "U.S." appears in your document, open the find and replace function in word (use CTRL + H) and replace "U.S." with "US". This will make it so that in step 2 carriage returns are only placed at the end of sentences.
    a. Other common terms that should be replaced: e.g., i.e., etc., vol., Jr.
    b. There are likely context specific terms in your transcript that you will also want to address.
2. In the "Find what" box, type "." and in the "Replace with" box type ". ^p". You can then either replace them one at a time using the "Replace" button or all at once using the "Replace All" button. If you use the "Replace All" button you will still likely need to review the entire document to make sure no sentences were separated incorrectly.
    a. If you have sentences that end in other punctuation marks (? or !) repeat step 2 using the other punctuations marks in the "Find what" box.

[2] We are using the term, "*Disease X"* to mean that this is a made up scenario and doesn't relate to any actual disease outbreak.

You can avoid this step entirely: if when you transcribe the data you build in carriage returns at the end of each sentence.

Regardless of which option you choose, now this example text will look like this:

Figure 3. Example of text after formatting by putting every sentence on its own line

Facilitator: What have you heard about *disease x*?
Respondent 1: Kid will have delays as they get older -- it causes problems for them.
And I worry about long term effects for adults too because no one seems to know about that.
Respondent 2: I learned yesterday that people can get reinfected even after having it one time.
Respondent 3: I didn't know that. You know, that kind of thing that makes me really scared.
I thought it was like the chicken pox.
Facilitator: Do you have any concerns or fears?
Respondent 1: I mean, at first, I did but now that I'm vaccinated and my kids are vaccinated?
And we just been home I mean, ninety percent of the time home alone only the kids go to school and we get groceries delivered so we don't a lot of risk.
Respondent 3: I'm less concerned about getting infected but about being treated for it.
Facilitator: What specifically concerns you about treatment?
Respondent 3: Well I know it's very hard to get a doctor's appointment and hospitals are incredibly busy so if we got sick where would we go?
And how do you know that you're sick enough to need treatment?
There isn't much information available about treatments for milder illness or how to recognize that mild symptoms are becoming severe.
Facilitator: And in regards to *disease X*, what you heard about like where did you gather all your information from?
Respondent 1: Every time I -- by the clinic.
I pick up the pamphlet and that stuff.

You will notice that every sentence now starts on its own line. When it is imported into Excel, each sentence will be on its own line. If your text has facilitator questions or other prompts that you would like to retain along with each statement, you can bold the question or prompt and add an additional carriage return ("enter") before each question or prompt. This is what the example text looks like after taking this optional step:

Figure 4. Example of text after formatting with questions or prompts

Facilitator: What have you heard about *disease X*?
Respondent 1: Kid will have delays as they get older -- it causes problems for them.
And I worry about long term effects for adults too because no one seems to know about that.
Respondent 2: I learned yesterday that people can get reinfected even after having it one time.
Respondent 3: I didn't know that. You know, that kind of thing that makes me really scared.
I thought it was like the chicken pox.

Facilitator: Do you have any concerns or fears?
Respondent 1: I mean, at first, I did but now that I'm vaccinated and my kids are vaccinated?
And we just been home I mean, ninety percent of the time home alone only the kids go to school and we get groceries delivered so we don't a lot of risk.
Respondent 3: I'm less concerned about getting infected but about being treated for it.

Facilitator: What specifically concerns you about treatment?

Respondent 3: Well I know it's very hard to get a doctor's appointment and hospitals are incredibly busy so if we got sick where would we go?

Facilitator: And how do you know that you're sick enough to need treatment?
There isn't much information available about treatments for milder illness or how to recognize that mild symptoms are becoming severe.

Facilitator: And in regards to *disease X*, what you heard about like where did you gather all your information from?
Respondent 1: Every time I -- by the clinic.
I pick up the pamphlet and that stuff.

Bolding the question or prompt will help you when you import the text into Excel because these questions or prompts will need to be moved into a different column from the text. Having them in bold will make it easier to see them to move them over to the cell just to the left of the text. Sometimes keeping the questions or prompts makes it easier to understand the text. This step is optional; you may wish to add these prompts or even section headings to your text, but it is not required.

Step 3. Paste the text into the Excel coding workspace – Once you have formatted your text so that it looks generally like ***either*** figure 3 or figure 4, you are ready to paste the text into the Excel tool. To paste in the text, you will open the **Coding workspace (5)** worksheet and paste your text into the "text sentences" column (column F). You should be able to paste just the first cell, and then the sentences will automatically fill in the column below, with one sentence in each cell.

You will notice that this column has some example rows of text already pasted in (rows 5 through 46). When you paste in your text, you should either delete these rows or paste your text over the examples. Since Excel automatically creates results tables that show the number and type of codes used, you will want to make sure that the examples are not left in the coding worksheet.

Take a minute to look at the three examples and then review the coding workspace worksheet rows 5 through 46 to look at how the rows have been filled. Example 1 (focus group) is in rows 5 through 20; example 2 (interview) is in rows 21 through 36; and example 3 (social media posts) is in rows 37 through 46.

You will also notice that column E, "Section Heading" has been filled in with focus group prompts, interviewer questions, and a social media post (that people were responding to) so that each coded text is next to the prompt or question associated with the sentence. Filling this column is not mandatory, but it can be helpful during your analysis because it will allow you to display the prompt next to the text and the code.

Step 4. (optional) Copy and paste the File ID number from the Transcript log sheet (2) worksheet – If you want to include the File ID number, the Participant type and the File name that you entered in the **Transcript log sheet (2)** worksheet, you can simply paste in the File ID number into all of the rows that you are coding. Once you've pasted the file ID column with the same ID from the Transcript Log Sheet table, the Participant type and the File name columns will automatically update using the information we've entered previously on the Transcript Log Sheet. This will be useful in the PivotTables section when you may be interested in stratifying your codes based on some demographic information.

4. Phase II. Coding the text

Once you have your text pasted into the coding workspace so that each sentence is a row, you are ready to begin coding.

Video tutorial 4 demonstrates how to code the example text using the epidemic coding scheme and the "action flags" coding scheme. video tutorial 4 Once you have viewed the viewed the video tutorial, you should be ready to try coding the example data.

Step 1. Review each sentence in your text and decide if you want to copy the entire sentence to the next column for coding. Typically, we code one sentence at a time. In the examples, you will also notice that in column G, the text from column F has been copied and pasted. The original text is placed in Column F. Before coding, copy the text from Column F to Column G. It's the text in Column G you'll use for coding. This allows you to either split or group sentences as needed.

For example, in row 10, you can see that the original text segment is, "I didn't know that." By itself, it is a difficult to understand what was meant. So, in column G, the analyst took that segment and added it to the previous statement, "I learned yesterday that people can get reinfected even after having it one time." You can see that cell G10 now contains a statement that is clearer and can be coded. You may not need to do any splitting or grouping of sentences, depending on what kind of text you have, but column G makes it possible to make these adjustments. It is recommended that you make these adjustments as part of the coding process, so you will decide what text is included in the cell, then assign the broad code, then the specific code, and then the action flag code (if you are using it), then proceed to the next row.

Step 2. Use the pre-formatted drop-down menus to assign codes to text segments – When you open the Coding workspace worksheet (5), you will see "Broad code" and "Specific code" and "Action flag" in columns H, I and J. Each of these columns are programmed so that when you click on the right side of the cell, a small downward pointing arrow appears. When you click on the arrow, a drop-down list of coding options appear that you can select from. This is where you will code your text.

Step 3. Double code when needed – You may decide that a sentence could be coded two different ways, because it represents two different ideas. If you want to do this, all you need to do is duplicate the row, and then enter a second code for the new row. The image below shows what a duplicated line with a different coding looks like.

Figure 5. Example of text sentences being duplicated and coded in two different ways

Text sentences	Text segments for coding	Broad code	Specific code
[F] What have you heard about Disease X?			
[R] 1: Kid will have delays as they get older -- it causes problems for them.	[R] 1: Kid will have delays as they get older -- it causes problems for them.	_5_About_experiences_of_the_ill	Survivors - long-term health consequences
[R] 1: Kid will have delays as they get older -- it causes problems for them.	[R] 1: Kid will have delays as they get older -- it causes problems for them.	_1_About_the_disease	Vulnerable groups to disease

Note that the duplicated row now becomes line #2. Make sure to update your line numbering (at the far left of the table, column A) so that each row has a unique number. This is called double coding.

Before going on to the next step, let's take a minute to become familiar with the rest of the columns in the **Coding workspace (5)** worksheet.

Columns and their purpose

A. **Line no:** sequentially numbered starting at 1 so that every row has a unique identifier.

B. **File name**: in this column you provide a name to each text that you import into the tool. Each listening text, notes set and set of social media posts will have its own name. Our three examples are "Disease_FG," "Disease_Int" and "FB_Post1". This allows you to identify which text set a coded sentence comes from. Once you enter these on this worksheet, the **Coding workspace (5)** worksheet will automatically update them based on the "File ID" number that you enter during the coding phase.

C. **Participant Type:** there is no rule about how to fill this in but make it something that is meaningful in the context of your analysis. For example, if you had focus groups of older women, younger women, and older men, then these would be good participant type names. As with File name, this can be made to automatically fill in the **Coding workspace (5)** worksheet by entering the File ID number.

D. **File ID:** this is a numeric file identifier that corresponds to the number you assigned to each file in column A of the Transcript log sheet.

E. **Section headings:** As explained above, this is an optional column for entering the question, prompt, or section header information that you might want linked to your text segment and code.

F. **Text sentences** users will paste their text sentences into this column (D)

G. **Text segments for coding:** you will copy the text from the "text sentences" column, and then either paste them as is, or split the sentence into smaller segments (or even merge them into multi-sentence segments) just before you code each line

H. **Broad code:** this column contains the drop-down list of broad topic codes.

I. **Specific code**: Once you select a broad topic code, a corresponding specific code list will appear. You will choose a specific code from that list.

J. **Action flags:** This optional coding scheme allows you to add a "flag" to text segments that you feel should receive immediate attention.

K. **Comments:** This column is not programmed, and you can enter any kind of text that you want. One useful approach is to suggest codes in the comments section when you code something as "other" because there is no appropriate existing code. If you find that you have multiple duplicate suggested codes in the "comments" column, that might indicate that you should add that term as a code to your coding scheme.

A note about coding technique – assigning codes to sentences is not an exact science. There will always be some subjectivity to understanding meaning of text segments, and two analysts will likely not code everything exactly the same way. When working as a team, the goal is usually to achieve consistency between coders, through a process of having two analysts code the same text and compare. Those codings which do not correspond can be discussed and resolved.

If you are working alone, or conducting a less formal analysis, you may decide to just code the data yourself, and when you are finished, share both the source texts and the themes that you identified with colleagues to solicit comments and other points of view.

Advanced user option - As you prepare your text for coding in the (5) Coding workspace worksheet, you may decide that you want to include other information with your sentences. For example, you may have notes or comments from external sources that you would want to note. You can add as many additional columns as you wish to categorize types of text sets. For example, you may want to note whether social media comments were sourced from Facebook, X [Twitter], or other website. Or if you have conducted multiple focus groups the location of the focus groups may be an important factor that you will want to be aware of during the interpretation of the data.

Step 4. Open pre-formatted pivot tables, refresh data and view tables and charts - With the text fully coded, you are now ready to view tables and charts that show which codes you used the most. The frequency is an indication of the prevalence of a particular topic throughout your text. It allows you get a sense of the relative frequency or rarity of certain ideas, since you will be able to see that some codes were used a lot, while others were not used at all. This video tutorial shows how to display your coded data in a pivot table. video tutorial 5

A. Once you've finished coding your text, **open the Coding tables (6)** worksheet. If the worksheet is zoomed out to 50%, you can see that there are two pivot tables and two bar charts on the worksheet. Make sure to "refresh" the data (go to "data" at the top, then select "refresh all.")

B. Zoom in to 100% and scroll over to the left so that you can see the "Broad topic codes" table and chart. If you have not entered any new text or coded any new data, these tables and charts are populated with the example (*disease x*) text.

The table shows the number of times a text segment was coded with each broad code, and the percentage of all "codings" (each time you code a segment, we call it a "coding") that are of a particular broad code. The bar chart shows the same information graphically.

C. You can also reorder the table so that the broad code with the highest frequency is at the top. To do this, **select the cell, C7 (the first percentage value), and right click on it.**

D. About half way down the drop down list that appears you will see "sort", which you will select.

E. Then select "sort largest to smallest." This will reorder your broad topic codes by frequency. It will also reorder the chart.

Now that you have seen the first pivot table and chart, we are going to look at how pivot tables work, so that you can customize this table, or any table you choose. You can change the coding schemes that are displayed in a table by first opening the "field list dialogue box."

Do this by right clicking anywhere in the table, which opens a dialogue box with "Show field List" option at the bottom. Selecting this option will pull up the "PivotTable Fields panel on the right side of the screen.

Figure 6. Opening the field list box using "Show Field List"

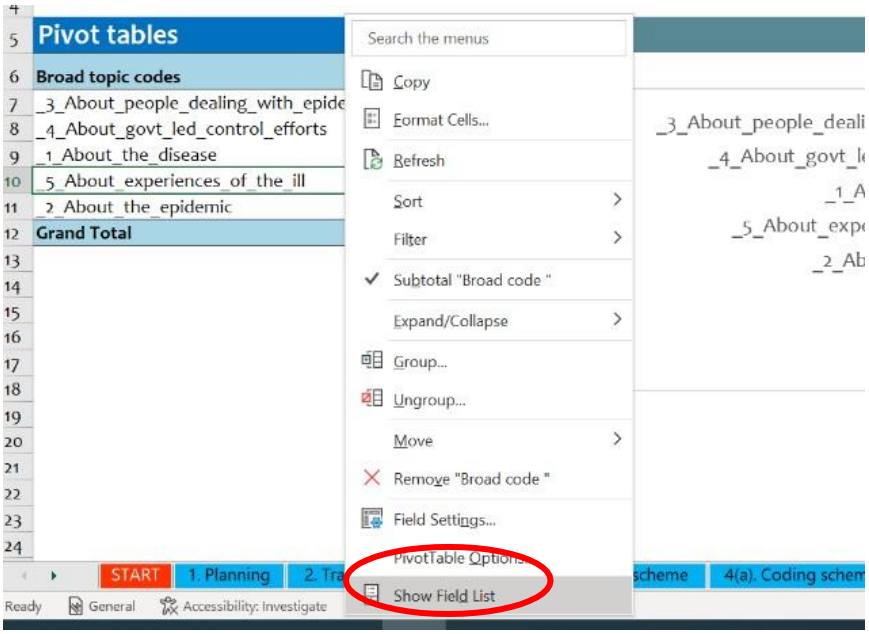
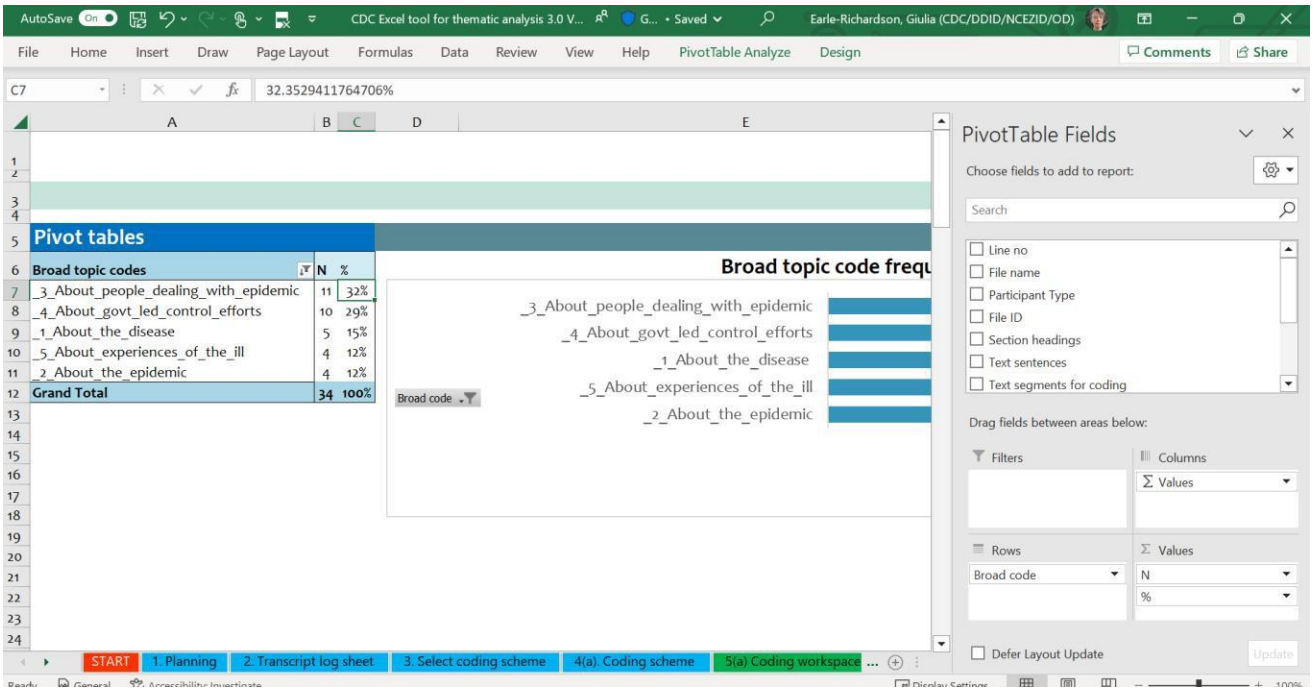


Figure 7 below shows what the pivot table looks like with the field list dialogue box open.

Figure 7. Pivot table with field list dialogue box open



The figure above shows the **Coding tables (6)** worksheet with the pivot table field list box open. It appears on the right side of the screen. The top box shows all of the names of the columns in the **Coding workspace (5)** worksheet that you might want to include in your table. You can see that for this table,

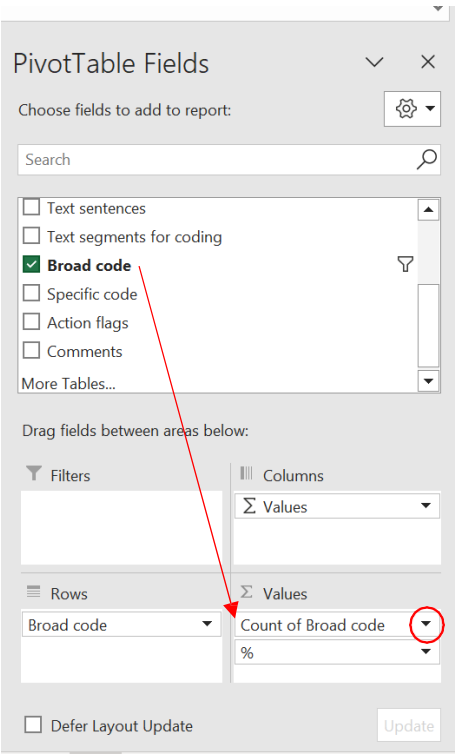
the “Broad code” column heading has been moved from the top list down into the box named “rows.” This tells Excel to make the pivot table display the Broad codes as **rows** in the table.
The four boxes below that box are places you can place the column names if you want to perform different operations. Very briefly, here are the operations performed when you place a column name in each box:

Filters: if you move a column name into this box, the entire table will be filtered by the values in that column. For example, if you want the table to show coded segments relating to government response efforts, you could drag “Broad topic code” into the ‘filter’ box. When you do this, another box opens with a list of Broad topic codes to select from. If you select one, for example, “About government led efforts to control spread,” then only the text segments (and their specific codes, etc.) that fit into that broad topic code would appear in your pivot table.

Columns: if you have columns that you want to present as columns you can put them in this box. Most often people use this when they have numeric data. You probably won’t use this much in this tool. But you can try moving a column heading into this box so you can see how it displays your text data.

Counts: this box will provide a numeric value for your text data. For example, we have dragged the “broad code” column heading down a second time (the first time we put it in rows), but this time placed it in the “Values” box. When you do this, the name of the heading automatically changes to “Count of Broad code,” meaning that in this box, the value being displayed will be a count. Figure 8 below shows what it looks like.

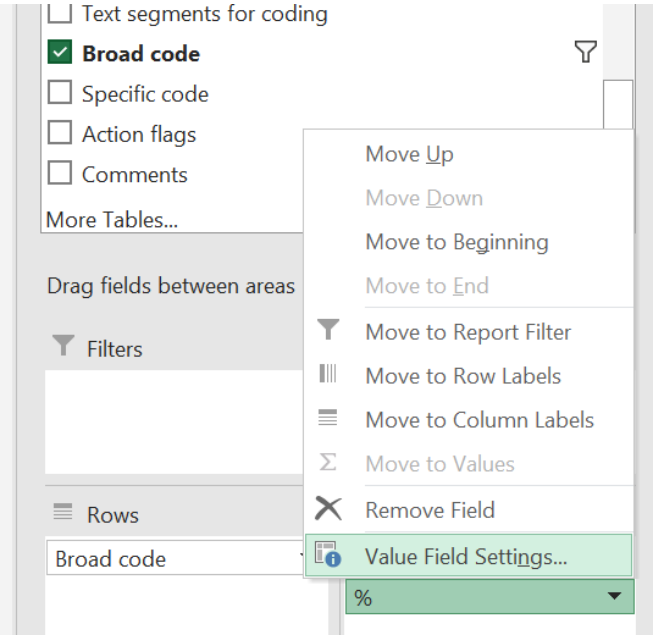
Figure 8a. Value field settings within “values” field setting box



This was done to make the pivot tables (see figure 7 above), however, once complete, the name of the column in the table was changed to “N” so that it would be easier to read. This change was then reflected in the pivot table fields box.

A similar process was used to create the “%” column, with one additional step. When the “Broad code” column name was dragged down into the Values box, it automatically switched to “County of Broad code,” just like the previous instance. However, if you select the downward pointing triangle to the right of the name (see red circle in Figure 8), a new drop down list appears, with “Value field settings” at the bottom. See example below, right.

Figure 8b. Value field settings option



Once the "Value Field settings" option is open, you will see that the default is "count." If you select the "show values as" tab, you will first see the screen shown in figure 9, showing that the default value for "Summarize values by" is "count." This can be left as it is. When you select the other tab, "Show Values As," the default is "no calculation." Left this way, the count will appear as a number. However, when we change this option to "% of Column Total" (figure 9.b.), it changes the display to a percentage.

Figure 9.a. Summarize values by…

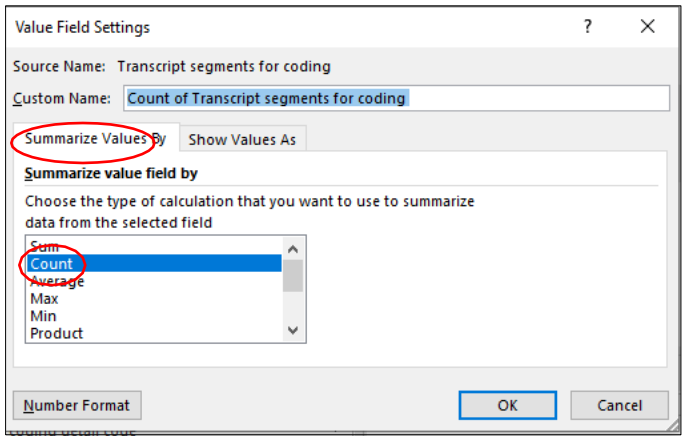
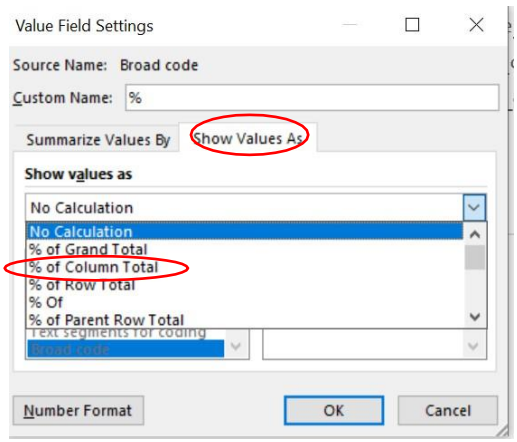


Figure 9.b. Show Values as…

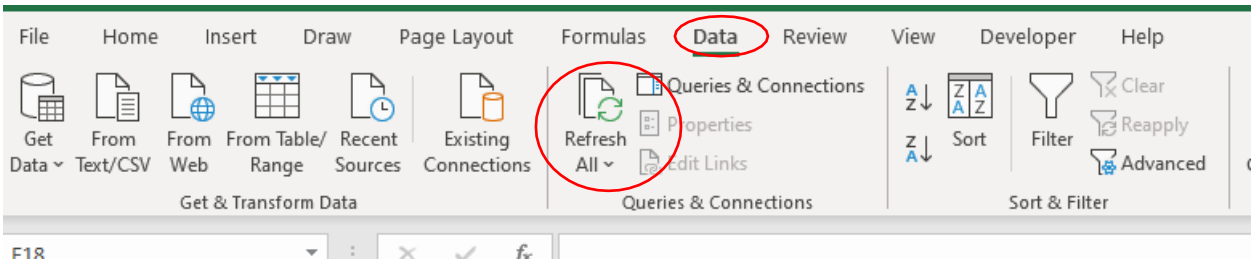


A really good way to learn how these boxes and settings options work is to try moving the column headings from the list to different boxes and see what the result is. In particular, experiment with putting multiple column headings in the "row" box. When multiple headings are there, they create a nested list, so that the lower items are subsets of the higher items. The best way to understand this is to change the ordering of the column headings within the "row" box and to look at the results.

Refresh your data whenever you modify your table or your data – It is important to remember when working with the pivot tables is that they do not update automatically. Any time you make changes to the information on the **Coding workspace (5)** worksheet, you should "refresh" the data to update the tables. You do this by going to the top ribbon, selecting "data" and then "refresh" and then "refresh all."

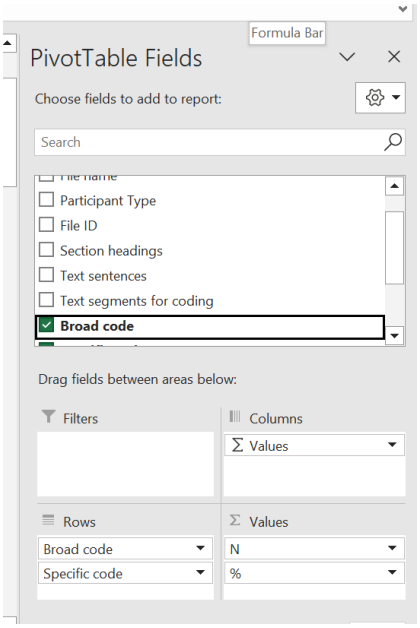
You can also refresh the data by clicking on the pivot tables that the "PivotTable Analyze" options appears on the ribbon and selecting "refresh" and then "Refresh all." The information in the pivot tables should now be updated to reflect your coded data. (See figure 6).

Figure 6. Command options for refreshing data



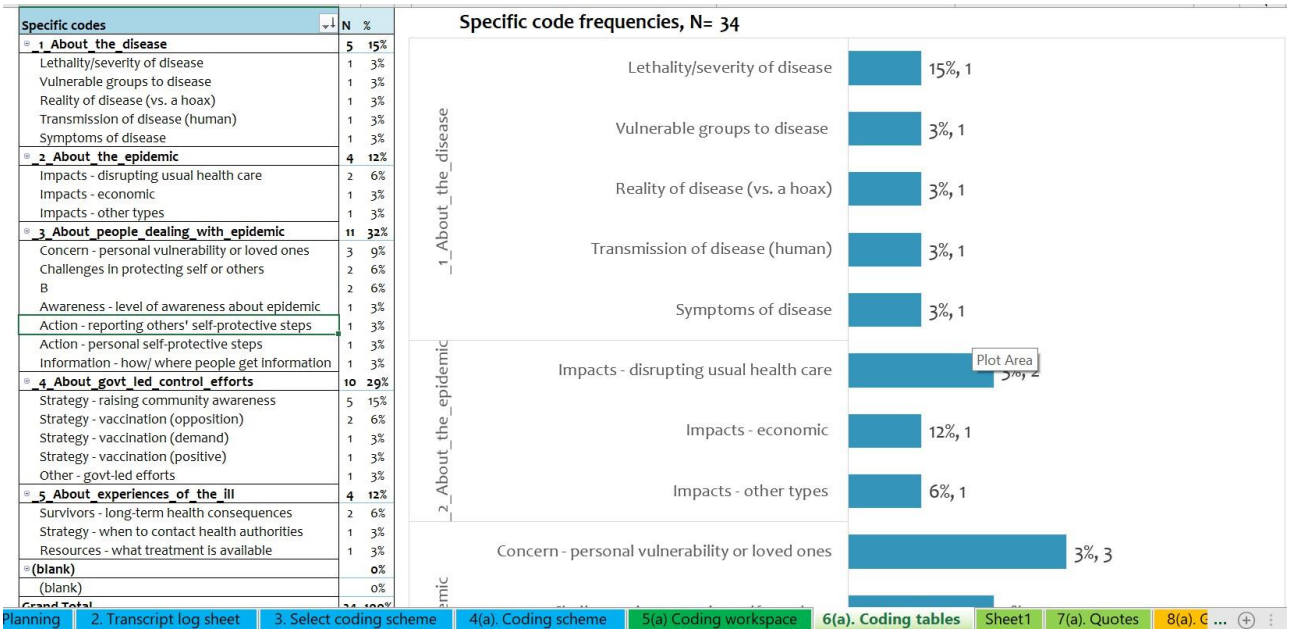
Getting back to the **Coding workspace (5)** worksheet, if you close all the boxes (field codes and field settings, and then scroll over to the right, you will see a second pivot table and chart set. If you were to open the "Show field list" box for this table (right click anywhere in the table, then go to the bottom of the list), you would see that "specific code" has been inserted into the "rows" box. This table displays the specific codes in frequency order, without respect to which Broad code a specific code is connected to. As a practice, try opening the "Show field list" option, then drag the column heading, "Broad code" down into the "rows" box, above the "Specific code" heading.

Figure 7a. Adding "Broad code" to Specific code table and charts



Look at how the table changes. In figure 7.b. below, we see how adding the "Broad code" heading creates a two-level table, with "Broad code" at the top and then the corresponding "specific codes" underneath. The chart also changes, with each set of bars of specific code frequency separated into a separate group according to Broad code.

Figure 7.b. Effect of adding "Broad code" to "rows" box in field codes boxes

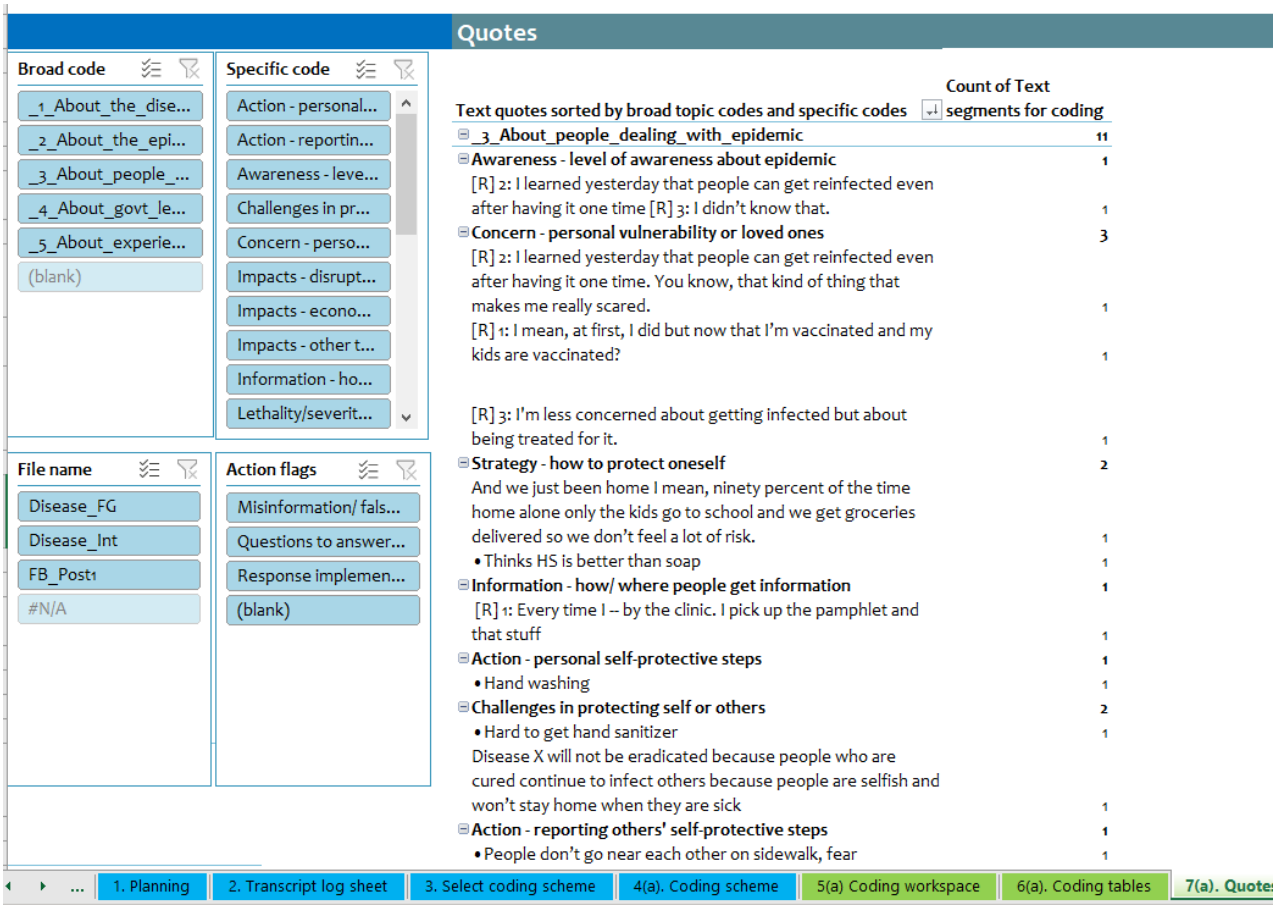


Step 5. Displaying important quotes – In addition to looking at the frequency distribution of your codes, it also can be very helpful to review the specific text segments assigned to each code. This can be helpful in a number of ways. First, it allows you to review the correspondence between the text segment and the code. You may have miscoded a segment, or you may decide that you would like to edit the code[3] slightly to better reflect the meaning of all the text segments as a group. If you are using the Immediate action flag coding scheme, looking at the quotes under each flag will quickly alert you to the issues you need to address right away. In addition, for your thematic analysis, you will want to find quotes that represent the themes and codes for your final write up. Using participants own words in expressing an idea is one of the most powerful aspects of this type of analysis.

Select the **Quotes (7)** worksheet. There are four "slicer" boxes that allow you to select part or all of your coded text in order to view the quotes.

[3] To edit a code, you would go back to the **Coding scheme (4)** worksheet, edit the code cell, then go back to the **Coding workspace (5)** worksheet, filter your coded rows so that only those with the old code are showing, and paste the new code from the drop-down list over all of the rows. As a last step, remove the filter so that all of the rows are then showing.

Figure 8. Four slicer boxes on the Quotes (7) worksheet (boxes on left)



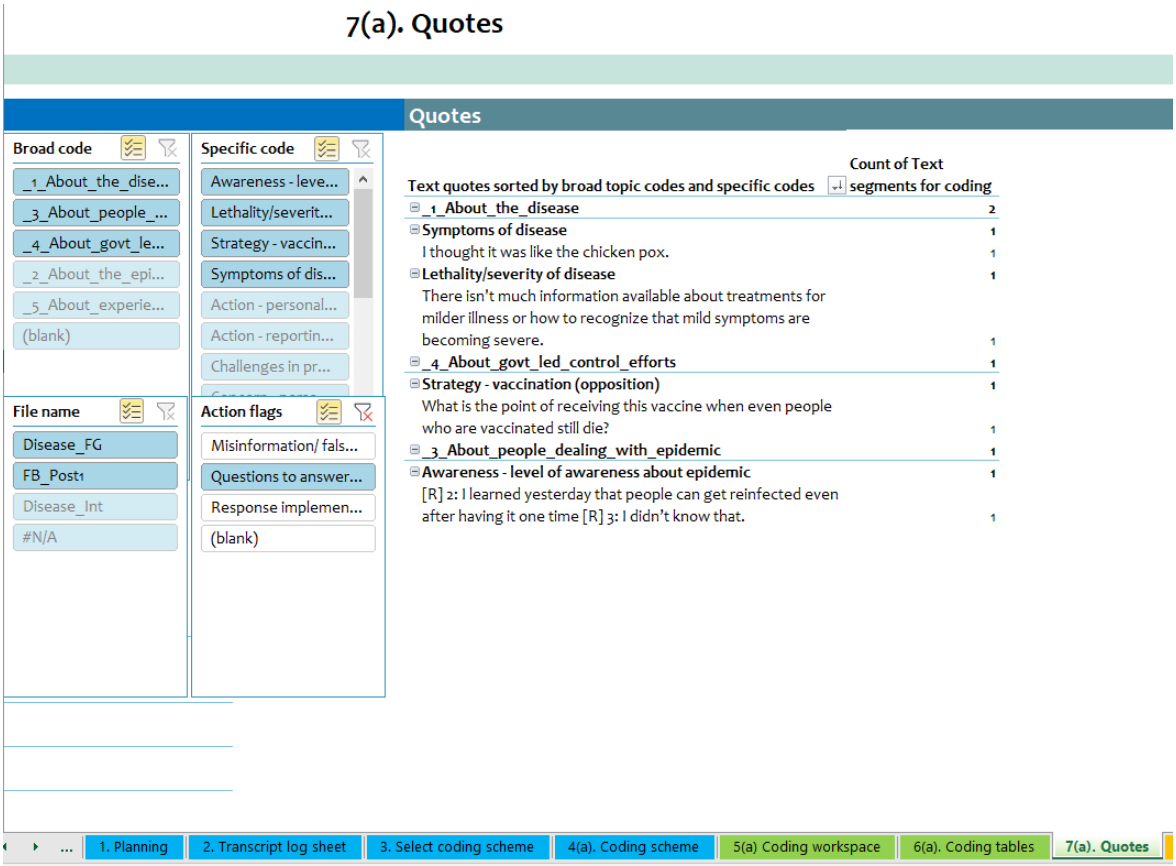
The default setting, shown above, is to show all of the Broad Codes and all of the Specific Codes. It also shows all of the files that were coded, and all of the quotes related to any "action flag". Put another way, all of the slicer filters are turned off. You can also turn off all of the filters (to show all of your quotes) by clicking on the "clear filter" symbol in the upper right-hand corner of each box. The symbol looks like this:  . To make sure that all your slicers are currently turned off, click once on the "clear filter" symbol in each of the four slicer boxes.

Your screen will look like figure 8 above. If your coding file isn't very large, or if you want to see all the quotes, this option will work fine for you. However, there are many situations where you will want to look at only quotes from one Broad code, one Specific code (within the Broad code), one file or one Action flag code. These slicers make it easy to narrow down the set of quotes to a particular group that you want to see. For example, suppose you wanted to see all of text that related to a question that a member of the public had. This could be very useful for creating a "Answers to commonly asked questions" document for public dissemination.

To display only quotes related to "questions to answer" code from "Action flag" coding scheme, go to the "Action flags" slicer box, and click once on "Misinformation/false information," "Response implementation," and "(blank)" buttons. This has the effect of deselecting all the buttons except "Questions to answer." This might seem counterintuitive, but if you remember that the buttons that are

shaded blue are the ones for which the quotes will show, the logic of deselecting all the buttons around the button you want to display makes sense. Figure 9 below shows what your worksheet looks like when you are filtering ONLY on "questions to answer" in the "Action flags" slicer box. In this box, "Questions to answer" is the only button that is shaded blue, and it is the only one whose quotes are displayed on the right.

Figure 9. Using slicer boxes to filter so that only text segments with the action flag, "questions to answer" are shown



Notice that in the other three slicer boxes, there are both dark blue shaded boxes with blank text, and lightly blue shaded boxes with grey text. What this tells you is that some of the Broad codes, specific codes and file names don't contain any quotes that were also coded as "Questions to answer" in the "Action flags" coding scheme. For example, in figure 9 above, you can see that only the Broad codes: "1. About the disease," "3. About people dealing with the epidemic," and "4. About government led control efforts" had quotes that were also coded in the "Action flags" coding scheme as "Questions to answer."

You can confirm this by looking at the quotes and seeing that they are all listed under one of these three broad codes.

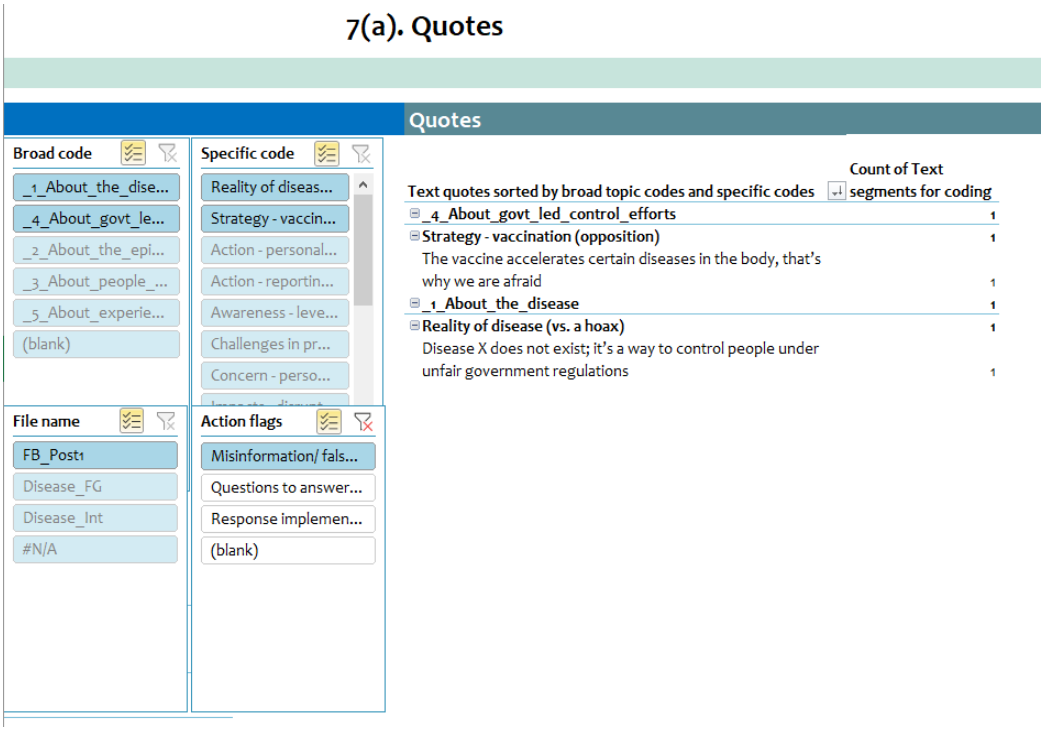
We recommend that when you are finished looking at quotes filtered a certain way, that you clear the filters on all four slicers for future use.

In most cases, you will select a button in only one of the slicers and leave the rest unfiltered. You typically will only select one button from that slicer. However, if you want to select more than one button from a slicer, you can select this button at the top of the slicer: 

This will make it so that you can select two or more buttons in any slicer. Just remember that buttons that are set to blue shading will be the quotes you see, and the buttons that are white will be the quotes that you *don't* see.

Try clearing the "Action flag" slicer box, and then clicking on the buttons until "questions to answer," "Response implementation," and "(blank)" are white, and "Misinformation/false information" is shaded blue. Your screen should look like this:

Figure 10. Quotes (7) worksheet showing quotes for "Misinformation/false information" action flag



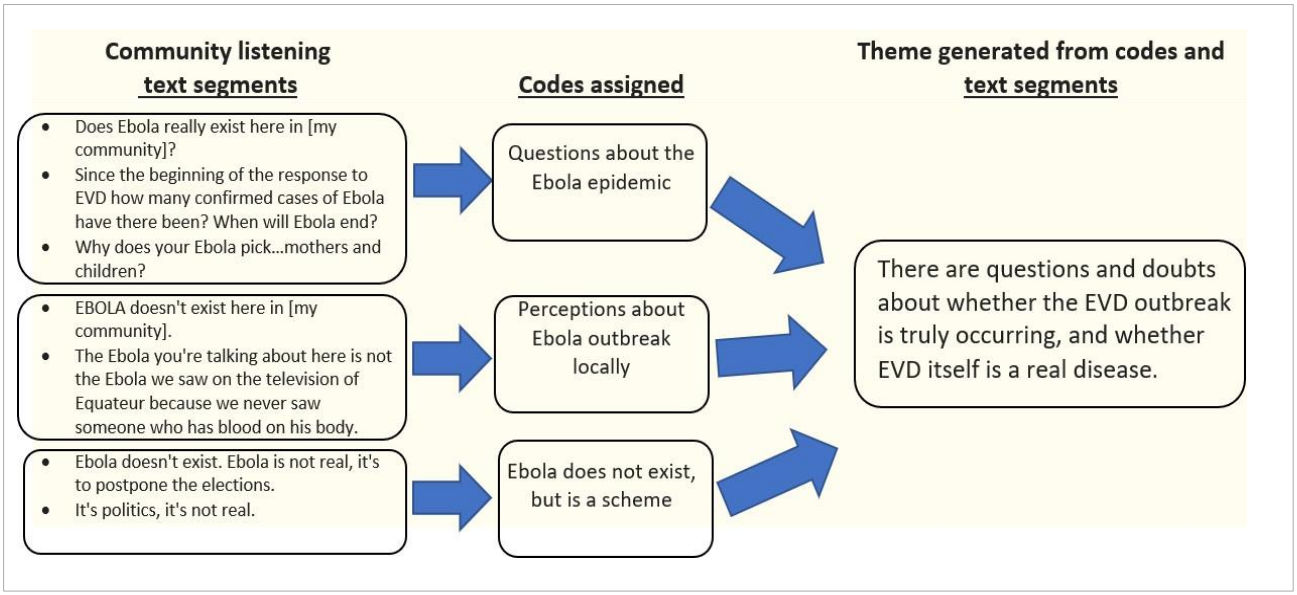
More information on PivotTables in Excel

5. Phase III. Generating themes from the text

Now that you have coded your text, and have viewed the codes that represent your text, you are ready to start building themes. In the context of emergency response, themes can best be thought of as statements about community perceptions, concerns, beliefs, sentiments, questions or behaviors that are suggested by combining codes used with your text. For example, figure 8 below is adapted from a thematic analysis of public attitudes toward Ebola virus disease in the Democratic Republic of Congo in 2018. While the codes themselves (Questions about the Ebola epidemic, Perceptions about Ebola outbreak locally, and Ebola does not exist, but is a scheme) are all different, when put together, they

suggest an overall theme of skepticism and doubt about the reality of the outbreak. This analysis was done as part of an analysis concerned with understanding why participation in Ebola response activities was low. This theme not only brings together the codes represented by the comments, but also relates to the overall inquiry about poor participation. While themes don't necessarily have to be written in full sentences, it can be helpful to do so just to be sure that there is enough detail so that the themes are informative for your response efforts. You want to avoid overly brief or vague themes, such as "doubts," "mistrust," or "concerns about the response."

Figure 11. Example of how text segments and their codes relate to a theme



Adapted from Earle-Richardson, et al., 2021.

It is also helpful to not that having the text segments next to the codes when developing the themes helps you understand much more fully what is meant by the code. For example, the code "questions about the Ebola epidemic" doesn't fully convey that some of the questions were in fact expressing doubt about the reality of Ebola. By looking at the text segments and codes together, you can pick up these nuances and make more representative themes. You may also decide to revise your code after you see all the text segments linked to it. This is a normal part of the analysis process.

Phase III: **Thematic analysis** – In this phase you will look at your text segments organized by code, and use them to develop themes. Depending on how many codings you have, you may decide to start with just one or a few of the most frequent broad topic codes. For the demonstration, we will take one broad code and all of the specific codes underneath (as well as the text segments underneath that). This way, you can see the process, and then decide if you want to do all of the text at once, or possibly break it into parts. A brief video tutorial reviewing this phase is available here: video tutorial 6

You will complete six steps:

Step 1. View the pivot table in the Grouping codes (8) worksheet and decide whether you want to analyze all of your text at once, or if you would rather just select a subset using certain Broad topic codes. We offer this choice because if you have a long text which has resulted in many coded text

segments, it can be a little overwhelming to work them into themes all at once. If you prefer, you can just select the most common Broad topic codes (in terms of the number of codings), or alternatively, the Broad topic codes that are most important to the questions you are trying to answer The slicer box in the upper left hand corner of the worksheet allows you to limit the pivot table display to only certain Broad topic codes. Remember, with a slicer, the codes that are shaded in blue are the ones that will appear in the pivot table. Those in white will not.

Open the **Grouping codes (8)** worksheet, and look at column A. For our example, "About the epidemic," and "About people dealing with the epidemic" have been selected to be included in the pivot table below for use in developing themes. You can always go back and do the process again with different codes.

Step 2. Copy each code and the text segments underneath it to one of the "idea boards" to the right. Columns E through K on the **Grouping codes (8)** worksheet contain four "idea boards" (boards A, B, C and D). When you first start, it won't matter which idea board you paste to, because the idea is to start putting codes (with their text segments always under them) next to others code that seem similar. In the example, the broad codes, "Challenges in protecting self or others," "Action - personal self-protective steps," "Strategy - how to protect oneself," and "Action - reporting others' self-protective steps" all were put on idea board A because the analyst felt that they were related. Once you have several similar codes on the board, try writing an initial brief theme that explains what the codes all have in common. When you are working with your own data, you can use the set of blank idea boards to the right of the example and the pivot table will only contain your coding (make sure you have deleted all the examples from the **Coding workspace (5)** worksheet and refreshed the data!)

Step 3. Open Building themes (9) worksheet and paste in your completed idea boards from the previous worksheet, one on top of the other, in column B. This vertical structure will make it possible to do the next thinking and writing steps, that need to go across the worksheet, left to right. Once ordered vertically, you proceed across the table, left to right, describing the meaning of each code in more detail, summarizing groups of codes and finally, synthesizing what you have written into a paragraph. This sheet also reminds you of your original questions (column G) so that your synthesis relates to your original aims.

Step 4. Follow the instructions at the top of columns C and D. This involves describing what each text segment means to you (in light of what you are interested in learning), (column C) then summarizing those interpretive statements into a few sentences to represent the entire group or theme. During this phase, you may decide that a certain text segment belongs better with a different code or a different theme. That is fine. Just move it and make a note of it. (See C12 as an example). The important thing is that the text segments begin to form a pattern that is informative. You may find that not all of the coded text is relevant to your questions. Once you have written a brief summary of each theme, look for one or two text segments that are really representative of the theme. Put them in the box below your description of the theme.

Step 5. Synthesize the coding frequency and the summaries into one coherent paragraph, revisiting your analytic goals that are shown in column G. In the example, you can see how the short theme summaries as well as the counts of different codes have been brought in to the final narrative. For this you will also refer to your original questions (column G). At this point in the thematic analysis, you may find it easier to write your thoughts in a word document rather than remaining in excel, but this is a

personal preference. This is the final product that you will share with colleagues, partners, and the community from whom you collected the information.

Step 6. Open the final worksheet, Results and action planning (10) worksheet and complete the rows – The first two columns ask you to apply your findings directly to your original questions and to consider what recommendations your findings suggest. Recommendations can include action as well as investigating new questions that have arisen (or questions that remain unanswered). The last three columns relate to devising a strategy for who you should share the information with, what action steps to consider, and how to share the results. Completing this worksheet helps you move from analysis to action.

6. Selected advanced user topics

1.1. Modifying the coding schemes

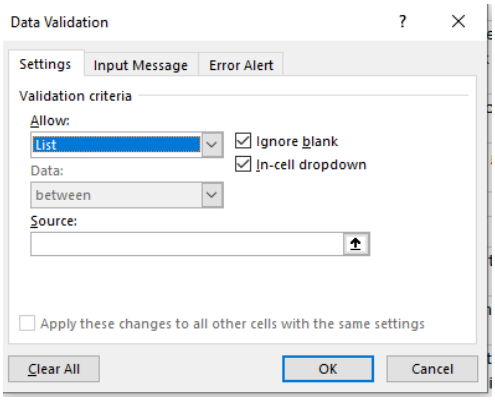
Depending on the nature of the data and type of project you are analyzing, you may need to add to or edit the coding scheme. You can add or modify any of the drop-down menus by going to whichever **Coding scheme (4)** worksheet you chose and editing the coding scheme. You can do this in two ways. You can type over an existing option or insert another specific code choice by right clicking on a cell and selecting "insert" and "shift down." This will create a blank cell that is included in the drop-down menu. When you make changes with either of these methods, the drop down menu on the **Coding workspace (5)** worksheet will change automatically. However, these strategies will only work for changing the Specific Codes and the Immediate Action Flag codes. Changing the Broad topic codes will affect the link between them and the specific codes. If you need to change a Broad Topic code, you should read section 1.3 Programming a two-level drop-down menu, below because you may need to reset the links between the Broad Topic codes and the specific codes.

Alternatively, if you need to make extensive modifications, you may want to simply use the blank coding scheme, and paste in the parts of the other coding schemes that you want to keep.

1.2. Programming the single level drop-down menus

The Broad topic area codes and the Immediate action codes are programmed using the "data validation" function (under "data" on the top ribbon). If you modify the coding scheme extensively, you may need to reformat the drop-down menus using this procedure. To program a single level drop-down menu, first select the range of cells for which you want to set a drop down list. Next, select "list" as shown in the box at the right, and then click once in the box under “Source.” Next go to the cells that you want included in the list and drag the mouse from the first cell to the last. This will set the drop down for that cell to be the range that you selected. Click on “OK.”

Figure 12. Illustration of the "data validation" dialogue box, when "list" is selected



1.3. Programming a two-level drop-down menu, where one depends on the other

This requires several steps.

1. **First, create a single level drop menu using the steps in 1.2.**
2. **Next, place that list on a worksheet in column A, making sure that these codes are formatted as shown in the figure below.** Notice that the entire code has been made into a single continuous text string by adding "_" symbols where you would typically have spaces. Please note that the Broad code can NOT begin with a number, include any special characters (including most accented letters), and has length limits. If you follow all of the directions below and you find that the dependent drop down menus are not working correctly, go back to your Broad codes and make sure that it begins with a letter or _, has no spaces or special characters, no accented letters, and is not long.

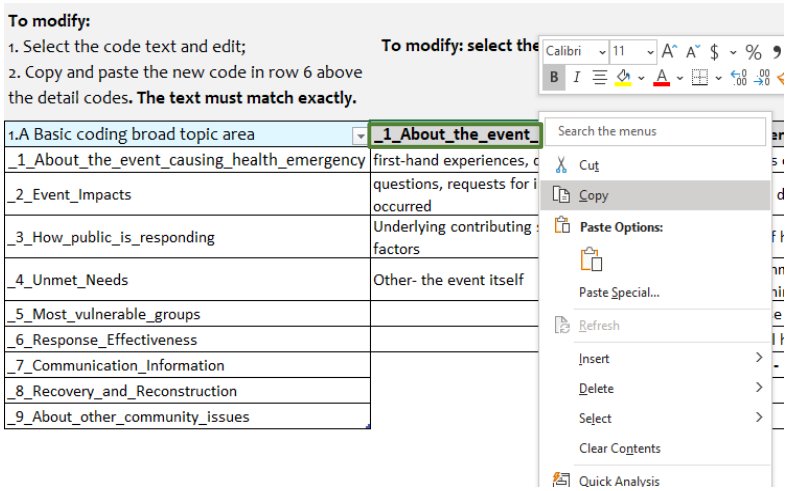
Figure 13. Formatted Broad code range and example of specific code (at right)

1.A Basic coding broad topic area	_1_About_the_event_causing_health_emerg
_1_About_the_event_causing_health_emergency	first-hand experiences, descriptions
_2_Event_Impacts	questions, requests for information about what occurred
_3_How_public_is_responding	Underlying contributing social context or other factors
_4_Unmet_Needs	Other- the event itself
_5_Most_vulnerable_groups	
_6_Response_Effectiveness	
_7_Communication_Information	
_8_Recovery_and_Reconstruction	
_9_About_other_community_issues	

3. **Next, place each of the broad codes horizontally along the top as shown above, so that you can list the contents of the specific codes underneath each one**. It is advisable to copy and paste the Broad codes from column A to the row across the top because the text needs to match exactly for the function to work.
4. **Once you have set up your worksheet with all the Broad codes listed vertically in column A, and then all the specific codes below the associated Broad code, you will name each range containing the specific code using the name of the Broad code.** To do this, there are two steps.

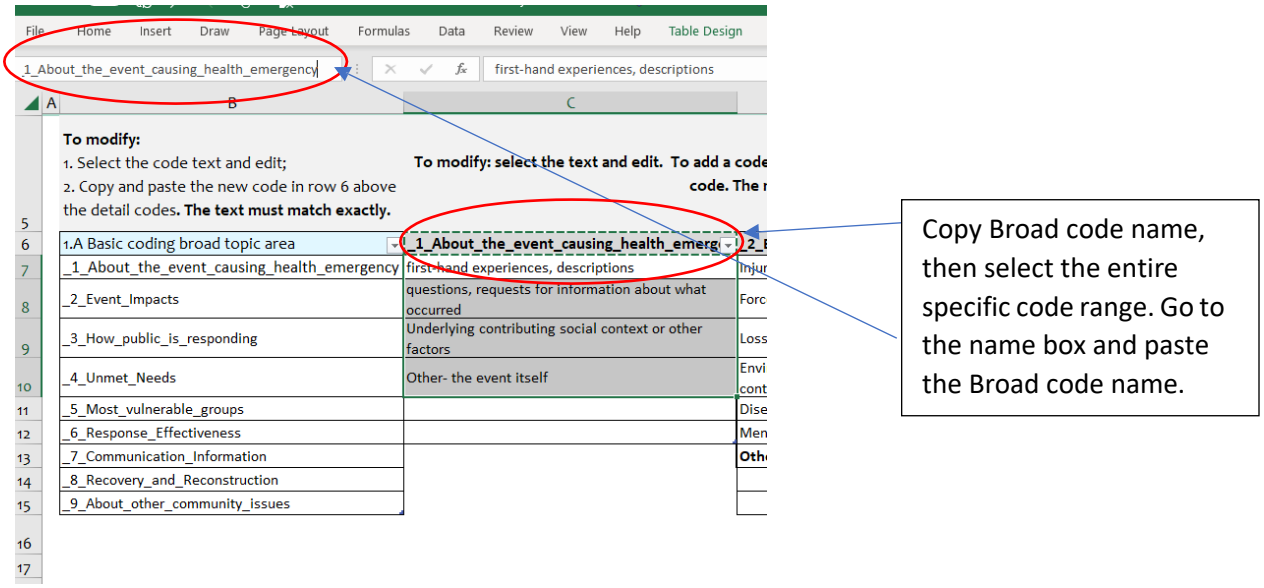
   4.a. Select the name of the first Broad code, and copy it, as shown in the image below.

Figure 14. Copying the name of the first Broad code



4.b. Then select the range underneath it (all of the specific codes) and go to the "name box" in the upper left corner and paste the name of the Broad code there.

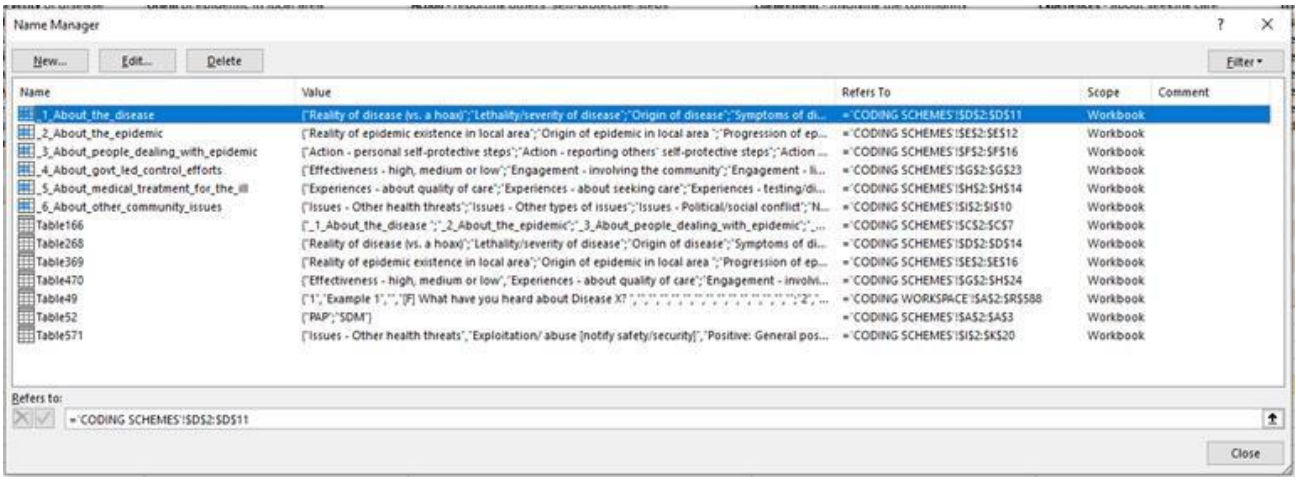
Figure 15. Naming ranges



Once you have named each of the specific code ranges using the corresponding Broad code range, you can proceed to the next step.

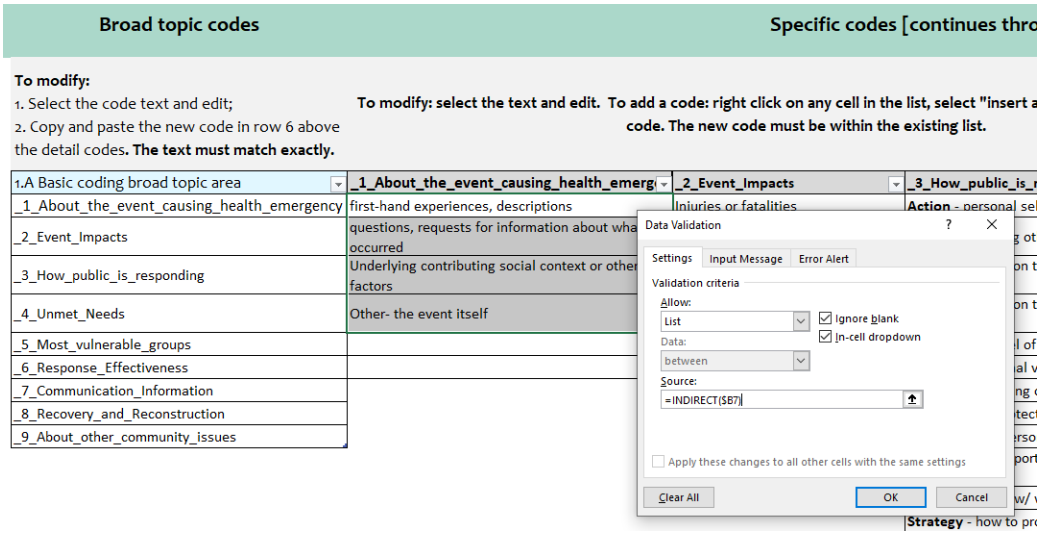
Alternatively – If you prefer, you can skip step 4a and 4b above and instead open a dialogue box to create and edit these named ranges using the "name manager." To find it, go to "Formulas," select "Defined Names," then select, "Name Manager." This dialogue box allows you to create, rename and edit named ranges, and can be easier to work with when you have a lot of named ranges.

Figure 16. Name manager box – use to add or edit named ranges



Step 5. Once you have named your ranges (using one or the other of the two methods above), you will need to open the 'data validation' dialogue box like you did for 1.2. Programming the single level drop-down menus. As before, you will start by selecting the range of cells for which you want to have the drop down menus appear. Ideally, this should be to the right of the Broad code cells, as shown in the figure below. In the validation drop down box, you will again select "list" from the validation criteria, but instead of entering a range of cells in "source," you will enter the following formula: =INDIRECT($[cell address of the Broad code]). This formula says to provide the drop down list that has the same name as the Broad code selected in the cell address that you provide. See the image below:

Figure 17. Programming a range of cells with drop down list (based on the selection of previous column) using data validation box



Note that the "indirect" function references cell B7, where the Broad code is. The "$" before the B tells the program to only look in column B. (If you were to insert other columns in between column B and C, the "$" would keep the reference column as B.)

Step 6. To check that this worked correctly, go to the Coding workspace (5) worksheet, select the Broad topic area that changed, and your new detail code should appear in the drop-down list of the Specific code column.

Now that you see how creating a two-level drop-down menu is made, you can modify the names of a Broad topic code by opening the "name manager" dialogue box and editing the name of the range to be the same as your new Broad code.

4.5. Working with multiple coders

It is very likely that while coding listening texts you will encounter situations where you want to have multiple coders coding the same data. There are several ways to do this, but the simplest is to have each coder start with their own copy of the file, segment the data and complete the coding separately and then copy the segments and codes from each coder into one file. The coders can then compare their codes, identify any discrepancies, discuss divergent codes and come to a final consensus.

Once coding is complete you will need to compare differences between how multiple coders may have applied codes to the data and come to final decisions about how to code the data. Generally, in an emergency response setting, you will focus on identifying and resolving coding inconsistencies; however, calculating an inter-rater reliability percentage (i.e., the percentage of codings that agree between two coders) is a good measure of how easy it is to use and clear your coding scheme. If you are going to use your coding scheme repeatedly, we recommend making this calculation and examining where areas of disagreement appear.

7. Appendix: Coding Schemes

Epidemics - Descriptive coding scheme

1 About the disease

Reality of disease (vs. a hoax)
Lethality/severity of disease
Origin of disease
Risk factors for contracting disease
Symptoms of disease
Transmission of disease (human)
Transmission to or from animals
Treatability of disease
Vulnerable groups to disease
Other - disease facts

2 About the epidemic

Compared to previous epidemics
Reality of epidemic existence in local area
Origin of epidemic in local area
Progression of epidemic in local area
Priority of epidemic vs. other issues
Underlying social factors
Impacts - disrupting usual health care
Impacts - social
Impacts - economic
Impacts - other types
Impacts - political
Other - epidemic characteristics

3 About people dealing with epidemic

Action - personal self-protective steps
Action - reporting others' self-protective steps
Action - opposition to response strategy (non-participation)
Action - opposition to response strategy (active opposition)
Awareness - level of awareness about epidemic
Concern - personal vulnerability or loved ones
Concern - reporting others' sense of concern
Challenges - in protecting self or others
Engagement - personal interaction with response
Engagement - reporting others' interaction with response
Information - how/ where people get information
Strategy - how to protect oneself
Strategy - how to protect others
Strategy - when self-protection is needed
Other - community response to outbreak

4 About govt led control efforts

Effectiveness - high
Effectiveness - medium
Effectiveness - low
Engagement - involving the community
Engagement - listening to community input
Motivation - people profiting from epidemic
Motivation - extending epidemic (not ending it)
Motivation - using epidemic for political gain
Motivation - intentionally harming people
Strategy - border control or checkpoint
Strategy - contact tracing / community surveillance
Strategy - isolation/quarantine
Strategy - legal mandated self-protection
Strategy - provision of masks/materials
Strategy - provision of water, soap
Strategy - raising community awareness
Strategy - testing
Strategy - vaccination (positive)
Strategy - vaccination (demand)
Strategy - vaccination (opposition)
Strategy - vaccination (neutral)
Strategy - burying the deceased
Strategy - other [add here]
Other - govt-led efforts

5 About experiences of the ill

Experiences - about quality of care
Experiences - about seeking care
Experiences - testing/diagnosis
Resources - where to go for treatment
Resources - where to go for testing
Resources - what treatment is available
Strategy - caring for the ill at home
Strategy - when to contact health authorities
Stigma - members of vulnerable populations
Survivors - long-term health consequences
Survivors - stigma, social consequences
Survivors - other needs
Other - treatment for the ill

6 About other community issues

Issues - Other health threats
Issues - Other types of issues
Issues - Political/social conflict
Needs - Economic assistance
Needs - Health infrastructure
Needs - Other types of needs
Needs - Safety and security
Needs - Water/sanitation

Epidemics - Action flags coding scheme

If you are coding in the midst of an outbreak, and you are in communication with health officials who can respond immediately to problems that come through in your analysis, you may want to use this coding scheme in addition to coding scheme #1. If you see text segments that indicate violence, exploitation, unintended effects of response activities, misinformation, questions, or other issues that the response would want to address immediately, this coding scheme helps you do that. The codes themselves set can be customized to the specific situation.

2. IMMEDIATE ACTION FLAGS
Exploitation/ abuse [notify safety/security]
Threats/ reports of violence [notify safety/security]
Misinformation/ false information [notify communications]
Questions to answer, information needs [notify communications]
Response implementation issue - case alerts
Response implementation issue - testing/diagnosis
Response implementation issue - medical care
Response implementation issue - contact tracing
Response implementation issue - vaccination
Response implementation issue - protective materials
Response implementation issue - other
Actionable suggestions
Other [specify]

Other Health Emergency Coding Scheme

1 About the event causing health emergency	Water/sanitation
first-hand experiences, descriptions	Protection and safety
questions, requests for information about what occurred	Access to other services
Underlying contributing social context or other factors	Other needs
Other- the event itself	**5 Most vulnerable groups**
2 Event Impacts	Protection of refugees and IDPs
Injuries or fatalities	Protection of children and women
Forced displacement	Other affected groups
Loss of homes and property	Other groups
Environmental destruction, contamination	**6 Response Effectiveness**
Disease outbreaks	Access to medical care
Mental health impacts	Access to other services and resources
Other - impacts	Timeliness of response
3 How public is responding	Quality of relief efforts
Action - personal self-protective steps	Coordination between agencies
Action - reporting others' self-protective steps	Resources allocated
Action - opposition to response strategies (non-participation)	Specific agencies or partners mentioned
Action - opposition to response strategies (active opposition)	**7 Communication Information**
Awareness - level of awareness about situation	Timeliness and accuracy of information
Concern - personal vulnerability or loved ones	Accessibility (languages, formats)
Concern - reporting others' sense of concern	Transparency of information
Challenges in protecting self or others	Channels: how transmitted
Engagement - personal interaction with response	**8 Recovery and Reconstruction**
Engagement - reporting others' interaction with response	Reconstruction and rebuilding efforts
Information - how/ where people get information	Support for affected businesses
Strategy - how to protect oneself	Economic recovery
Strategy - how to protect others	**9 About other community issues**
Strategy - when self-protection is needed	Political/social issues
Other - community response to outbreak	Economic issues
4 Unmet Needs	Health care infrastructure
Medical assistance	Other health threats
Shelter	Other types of issues
Food	Needs - Other types of needs

Emotional tone coding scheme

If you are interested in understanding the emotional tone of the text or the community comments, this coding scheme can also be used. This is not currently built into the tool, and you would need to add it to your coding workspace. This type of coding can be very useful if you are trying to understand whether community members are feeling positively or negatively about response activities, whether they are feeling afraid, discouraged, or hopeful about the outbreak. It can be useful to track these sentiments in conjunction with the descriptive themes revealed by using Coding Scheme #1 through the course of the outbreak.

Emotional tone coding scheme
Positive: General positive
Positive: Gratitude
Positive: Happiness
Positive: Optimism
Positive: Pride
Positive: Unity
Neutral: Indifference
Neutral: No particular feeling
Neutral: Surprise
Negative: Sorrow
Negative: Anger
Negative: Concern (or worry)
Negative: Confusion
Negative: Despair
Negative: Disappointment
Negative: Doubt/skepticism
Negative: Fear
Negative: Mistrust/suspicion
No emotion can be identified