SAGE research**methods**
datasets

Learn to Build a Codebook for a Generic Qualitative Study

Learn to Build a Codebook for a Generic Qualitative Study

Student Guide

Introduction

Coding is a common technique in qualitative analysis for condensing data into identifiable topics. A code is a distilled topic applied to a text segment illustrating that topic. By using codes, researchers are able to search for topics across data and thereby identify patterns and what stands out from patterns. In addition to creating codes, researchers create a codebook for any given study; a codebook is a list of codes with code definitions, allowing researchers to keep track of how codes are being used to make sense of data. In some cases, code definitions come from the literature review. In other cases, codes are defined by how participants refer to a topic. For example, a survivor of intimate partner violence may define "control" differently from how it is defined in a literature review. Hence, researchers must decide whether their study will be primarily driven by a literature review or by the data themselves. Codebooks are also products of research, illustrating what topics are relevant for a particular research question.

What Is a Codebook?

Qualitative analysis uses several techniques to keep track of critical topics in the data. These can include memo writing, diagrams, and coding. Coding data is a generative process that focuses on a close reading of data in order to capture, as best as possible, participant assumptions, insights, and complex motivations. This phase is sometimes called "first-cycle" coding where we apply many condensed topics or "codes" to small pieces of data (Saldaña, 2016). This process may focus

on various contextual units—sentences, paragraphs, or textual segments of other sizes. Early codes may be either deductive or inductive. Deductive or *a priori* codes come from a theoretical or conceptual framework, whereas inductive codes come from new ideas we discern in the data themselves. If codes are adopted from a framework, it is nonetheless important to define and describe how they will be applied to data. Codes that are "emerging" or inductive are constructed as researchers review data.

Codes attached to parts are themselves partial, but after generating a substantial number of codes, they are reassessed, renamed, and merged with other codes. Because it is critical to understand how a code is to be applied, researchers use code definitions to operationalize these topics. A list of codes with their definitions is called a "codebook." In basic qualitative studies, there is no agreed upon approach to what a definition should contain, but the following components can be used: description/definition, origin, importance, example, counterexample, and reflection (Maietta, Hamilton, Swartout, & Petruzzelli, 2018). A **description** or definition of a code offers guidance regarding its application. For example, the code *self-growth* might be defined as any text referring to an incident leading to new skills, understanding, or self-concept. The **origin** of the code refers to when it first became part of the study. It may have been evident during the literature review, or the researcher might have become aware of it much later. Keeping track of where codes come from can be useful later when making claims about a deductive or inductive process. The origins of codes will make it clear whether the topics in a study were primarily *a priori* or whether they emerged during data review. The **importance** of a code is the rationale for why it matters to the study. In a study of life satisfaction, we might employ a code for *friendship*, but we probably do not need numerous other codes regarding friendship. An **example** of a code is a verbatim quotation to which the code was applied. A **counterexample** (or anti-example) is a text segment that might be confused for the code but which is not given the definition. For example, we might decide not to code a text segment as

Learn to Build a Codebook for a Generic Qualitative Study

self-growth if the text refers to someone's promotion at work but does not suggest how this influenced their self-concept or skill building. Hence, examples and anti-examples essentially refer to inclusion and exclusion criteria. **Reflection** refers to how the code has evolved as more data are coded. For example, we might notice that a code for *control* in reference to survivors of intimate partner violence includes references to endurance because some participants discuss endurance as a form of personal control of one's circumstances. In the code definition, we might also clarify whether the code is descriptive, capturing a surface-level description, or interpretive, a more abstract topic capturing subtext and inferred meaning. A topic such as *emotions* would generally be descriptive, whereas a code such as *emotional labor* would be more interpretive based on an abstract conceptualization of data.

In investigating a code, we might also use memos to elaborate on its properties and dimensions. For example, a code for *victim* might have the property *visibility* which has the dimension *high* to *low*. A victim might be open about their victim status or furtive about it. A rich description of a code (containing properties and dimensions) might be used instead of creating numerous subcodes. Deciding whether a code should have subcodes depends on the nature of the research question and whether you would like to easily review and have access to those subcode topics in reports and present them in diagrams.

In a generic qualitative study, we might begin with a start list of codes—a point of departure based on a literature review or disciplinary topics of concern—but also identify and construct inductive codes (Saldaña, 2016). A study that employs both deductive and inductive codes invites researchers to carry with them what they know from the literature and also listen to what is "new" in the data, possibly problematizing earlier assumptions or logics. In this case, we might find that concepts in the literature are interpreted differently by the participants themselves. A cancer survivor who calls himself a "cancer graduate" might invite us to rethink

the medical definition of "survivor." In data-driven studies, we might also pay close attention to *in vivo* codes, codes based on a word or a phrase spoken by a participant, such as a hurricane survivor talking about the event as a "war zone" or "chaos."

A codebook is not meant to be an exhaustive list of all topics that a researcher will ultimately write about in their analysis. Codes can be combined to construct more abstract concepts that we call "themes" because they are more evocative or abstract. A researcher might, for example, analyze the co-occurrence of codes to generate themes. The co-occurrence of *dream* and *decision making* might lead to the theme *ruling out the dream*. Hence, *ruling out the dream* would not be part of the codebook *per se* but would be developed into a theme and written about in the study's results. Having a solid codebook to draw from makes the process of constructing themes more efficient and rigorous.

Illustrative Example

This dataset example uses narratives written or typed by undergraduate students attending East Carolina University in 1999. As narrative analysts, we are interested in differences in how the narrators tell their stories and what these stories reveal about living through a hurricane and its aftermath.

> **Research question:** How do individuals experience living through a natural disaster?

The Data

The 42 documents, handwritten or typed, were collected at East Carolina University (North Carolina, the United States), in a history class of undergraduate students who were asked to write about their experience of the hurricane. In this approach to data collection, the researcher has less opportunity to ask for

clarification or for participants to expand upon elements of their experience, but the advantage is that each participant focused on what mattered most for them. They told their story on their own terms with no probing from the researchers. Some students chose to address a particular person as their reader and used an epistolary mode in narration. Others used dates to indicate how circumstances changed from day to day. All narratives were entered into a text analysis software for coding and analysis. Software allows researchers to apply codes to text and to generate quotation reports for each code.

Building a Codebook

Our research question is about experience: How do people experience living through a hurricane? We are also interested in topics related to experience, such as people and settings. Hence, when we began this study, we had a small set of *a priori* (or deductive) codes. These were based on assumptions regarding what young adults would write about in their hurricane accounts. The deductive codes were *emotions*, *family*, *friends*, *neighbors*, *decision making*, *media*, *flood conditions*, and *property loss*. Defining deductive codes is critical because it documents our point of departure and which codes were part of our understanding perhaps even before data collection or data review. When we also code inductively, we are able to discern how reviewing the data took us in a different direction from our predetermined topics.

Deciding to what degree we will use deductive and inductive codes is beyond the scope of this dataset, but we will point out that in studies driven by a theoretical or conceptual framework, we might exclusively use deductive codes. In studies that are more exploratory or data driven, we would instead incorporate a large number of inductive codes or perhaps only inductive codes.

The inductive codes we created for this project include *action*, *timescale*, *authority*, *metaphors, socializing*, *assessing physical environment*, *negation, speculation*,

sarcasm, and *absolutes*. However, we decided that some of these codes did not make it onto our final code list; other codes were renamed to better capture the topic. For example, *speculation*, *metaphors*, and *sarcasm* did not show up often in the data for us to consider them useful. Each of these was applied no more than three times in over 40 documents. Hence, *sarcasm* was merged with *negative emotions* rather than deleted. In contrast, the code *negation* was applied frequently, but we decided to remove this code from our list because we it did not adequately contribute to addressing the research question. Because this code captured the use of negative language such as "cannot," "didn't," and "wouldn't," it is better suited for a discursive study regarding how language is used to construct a discourse around a particular event. Also, for clarification, *absolutes* was renamed as *absolutist language*. *Loss* was renamed *experiencing loss* to emphasize the larger process it suggests; as such, it became a subcode of the larger code, *actions*.

Our approach, both deductive and inductive, was pragmatic. In other words, we were interested in coding for the *who*, *what*, *where*, and *when* of the data. Our codes address each of these fundamental domains. The *who* was addressed by the people codes, the *where* the *environment* codes, the *when* the time code, and the *what* the *person*, *communication*, and *emotions* codes. Codes such as *person* and *communication* serve as placeholder or organizing codes under which other codes are placed.

Here is the structure of the combined deductive and inductive codes.

Person:

- absolutist language
- beliefs
- religion

Communication:

- media
- messaging

Actions:

- assessing environment
- decision making
- experiencing loss
- helping others
- socializing
- trying to live normally

Emotions:

- compassion
- lack of control/agency
- gratitude
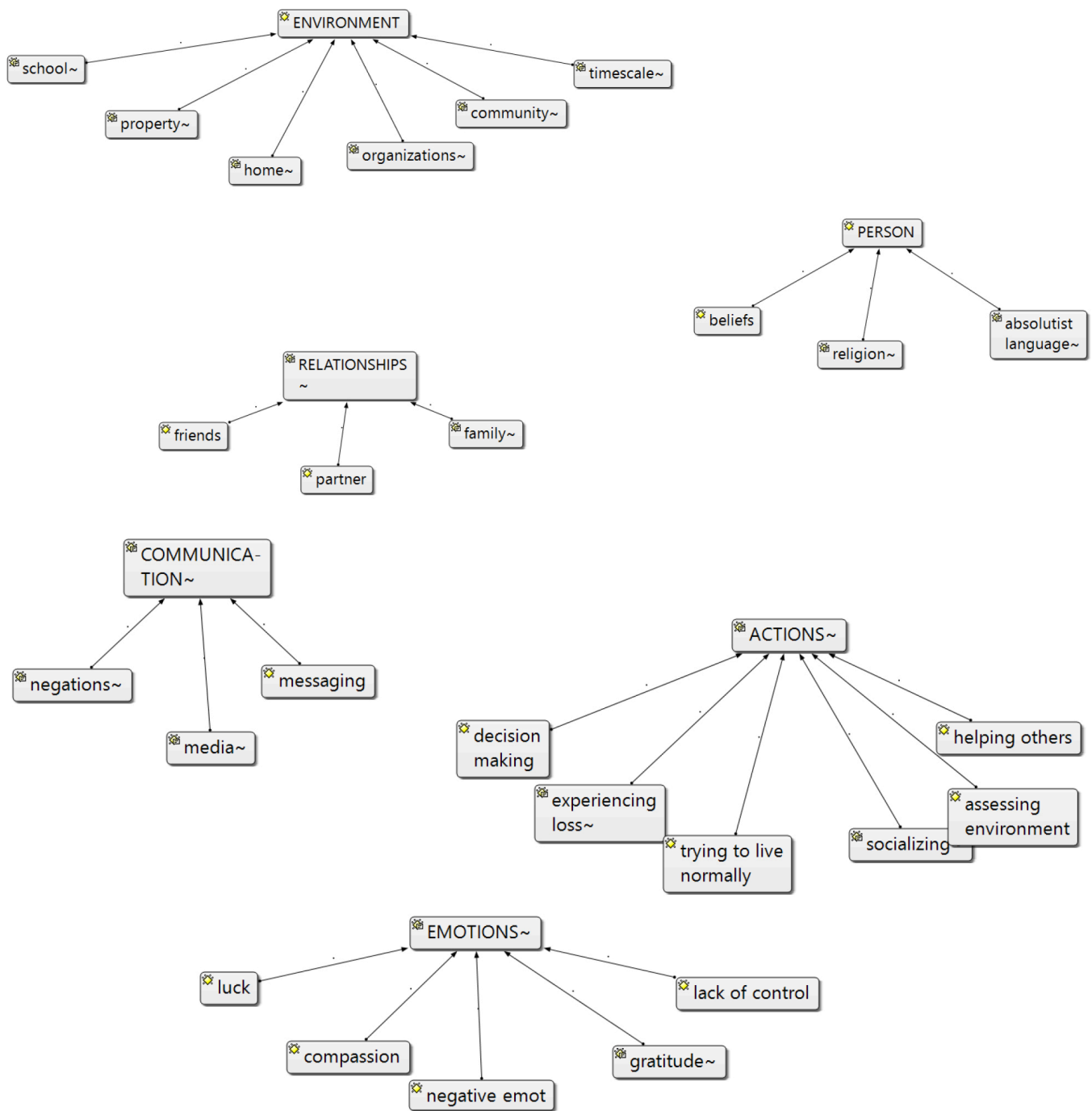- luck
- negative emotions

Environment:

- community
- home
- organizations
- property
- school
- timescale

People:

- family
- friends
- partner

Code Diagram

Below is a diagram showing the hierarchical connections among codes. Note that *lack of control* is not an emotion *per se*, but we linked it to *emotions* because of its close ties to one's emotional state.

ENVIRONMENT

school~

property~

home~

organizations~

community~

timescale~

PERSON

beliefs

religion~

absolutist
language~

RELATIONSHIPS
~

friends

partner

family~

COMMUNICA-
TION~

negations~

media~

messaging

ACTIONS~

decision
making

experiencing
loss~

trying to live
normally

socializing

helping others

assessing
environment

EMOTIONS~

luck

compassion

negative emot

gratitude~

lack of control

Example of Coding

Below is an example of how codes were applied to excerpts from Gina's
document. Notice that we tended to code entire paragraphs. The contextual unit
used depends on the researcher's objectives and the data themselves. In these
documents, participants were the ones who decided on the length of paragraphs.

Hence, we thought it would be useful to take advantage of the hurricane survivors' own sense of contextual segments as we applied codes. Notice also that, in general, several codes were applied to the same text segment in order to capture all relevant (and overlapping) topics.

Gina's document	
I f I could only use one word to describe the flood of 1999, it would be "overwhelming." I lived in Burlington, North Carolina my entire life and I have **never** experienced a disaster that I could come close to comparing with this**. I did not personally lose any possessions throughout this flood but it affected me a great deal.**	absolutist language experiencing loss
The one thing I did lose was the small amount of structure that I had in my life. Since I moved to Greenville in 1997, I have tried to gain some sort of stability. It has been a struggle trying to figure out who I am and what I am here for. Last semester I decided to declare my major, elementary education. Since then I have been taking school seriously and working on improving my grades. This semester I have been working extremely hard to keep my grades up while maintaining a stressful job waitressing.	experiencing loss
I started following the hurricane and realized that it was going to be serious. But I had to work that night so I did not have a chance to get out of town. I stayed at a **friend's** house that night and slept through the hurricane. I **couldn't understand** why we had a curfew until we talked to **Tar River Estates** and saw the flooding. I have seen **pictures on the news** of flooding in other areas but **you never understand until you see it in person.**	friends community communication media relationships absolutist language
I cannot even identify how this has affected me emotionally and mentally. The stress of not knowing what was going on with school or work was **very hard on me**. **Now, I am reluctant to drive or go out in the rain**. I have a few **friends** that **lost their homes** and I cannot begin to imagine what is going through their heads. **It is scary** that one event that **no one can** control can devastate so many people. When I see **the news stories** of families that have lost everything, I wonder how they will ever recover because the majority of them were barely scraping by before the flooding.	emotions experiencing loss media timescale friends lack of control trying to live

	normally

In the final paragraph, we applied the code for *emotions* because Gina indicates that not knowing what was going on was "hard on me" and that these circumstances are "scary." *Experiencing loss* was applied because Gina refers to friends who "lost their homes" and families who have "lost everything." The code *media* was applied because of references to news stories. *Timescale* was applied because of references to "now" and "when," which provide information on how Gina is experiencing time. *Friends* was applied because of the explicit mention of friends. *Lack of control* was applied because Gina refers to not knowing what is "going on" and being unable to "identify how this has affected me." The code *actions* was not explicitly applied, but because experiencing loss is a subcode of actions, the latter code is implied. Note that even though Gina mentions "families," we did not apply the code for *family* because she is not referencing *her* family. This is the kind of distinction we should make clear in the definition for this code. How a code is defined depends on our intentions for the study. Because we are interested in the experience of living through a hurricane, the more narrow application of *family* will help reveal how survivors experience their own family differently from their friends and romantic partners.

Notice that some codes are strictly descriptive, such as *friends*. Descriptive codes are topics that are lower on the conceptual ladder than interpretive or more abstract codes. Discerning an interpretive code, such as *lack of control*, might require returning to data more than once. For example, in the quotation below, Gina describes her place of employment, a restaurant that was destroyed, and how the world of the restaurant is now gone forever. Her account of this loss suggests a lack of control or agency. There's a finality in how she expresses it; she does not seem to have any control in light of the enormity of the loss. She does not explicitly say that she feels out of control, but we can argue that the quotation implicitly suggests the absence of personal power.

The Darryl's that I work at has to be closed down due to the flood. I walked through the restaurant after the waters had receded and cold not believe what Floyd had done. I immediately started to cry. Table and chairs were moved around or tossed over, everything had a nasty film on it, and the kitchen was destroyed. It was also sad to think that I would never be returning to see all the familiar faces and friends that I had made. (Gina)

In studies that attempt to construct themes or build theory, there could be numerous interpretive codes. In studies that are more descriptive, we might focus more surface-level codes. Interpretive codes require more elaborate definitions to ensure that we understand their conceptual nature.

Codebook

Below are excerpts from the study's codebook. Note that not all codes require a counterexample. A counter- or anti-example would only be necessary to clarify when a code should not be applied. For example, in the above data example, we did not apply *timescale* to the paragraph regarding how Gina's sense of how time passes in school (e.g., "Last semester I decided to declare my major.") because this has nothing to do with the hurricane. Also, a code reflection has been included for some codes and not others because we are still in the early stages of analysis. As we apply codes to more data, we will continue elaborating this part of the definitions.

Person

Definition: *Person* is a placeholder code used to organize subcodes underneath it. It will not be applied to text *per se* but be used to organize related codes regarding topics of interest at the level of the individual, such as beliefs.
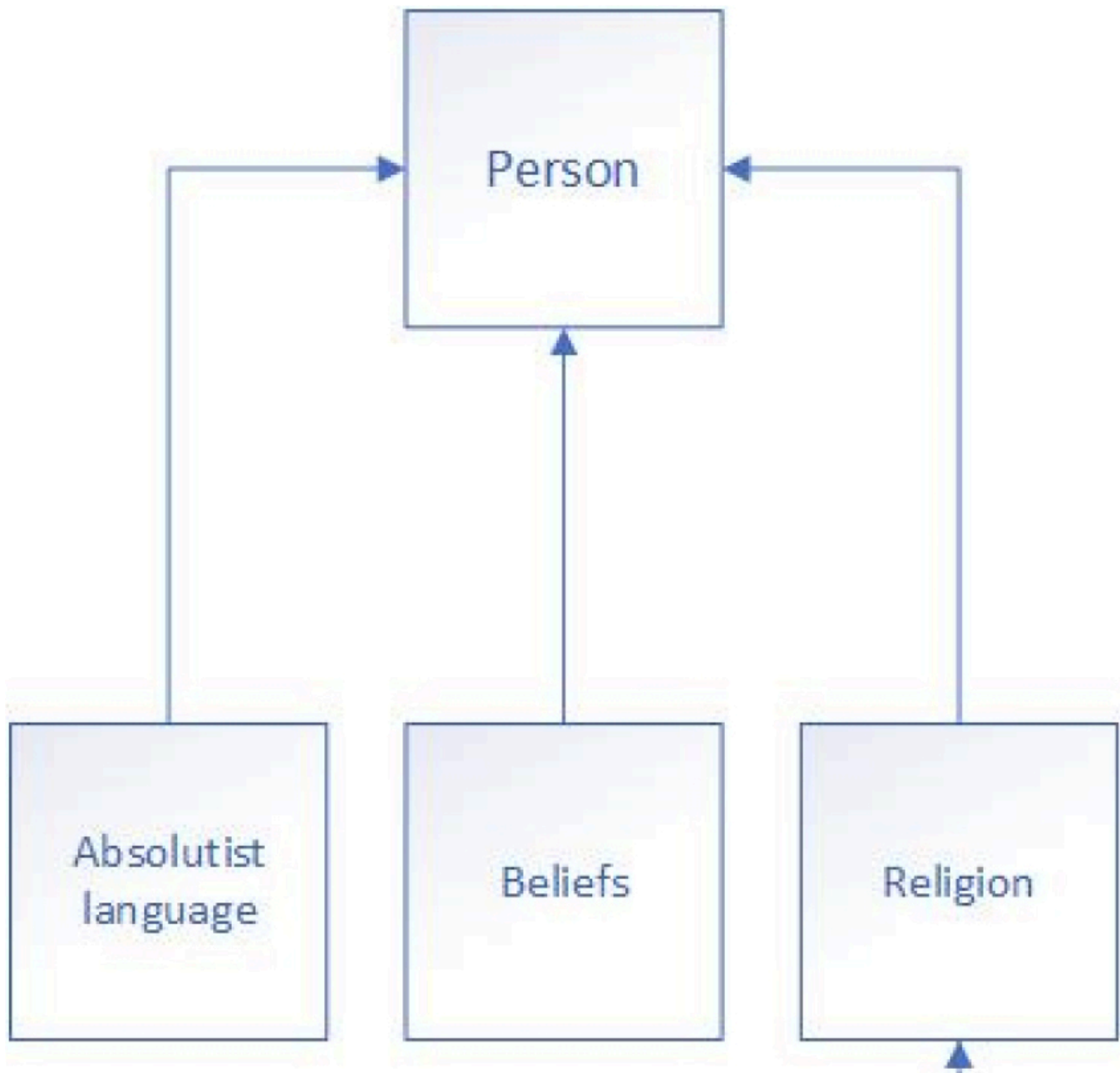
Absolutist language

Definition: The code *absolutist language* refers to language suggesting absolute magnitudes such as:

absolutely
always
completely

constantly

definitely

entire

extremely

ever

every

everyone

everything

full

must

never

nothing

totally

unbelievable

These words point to extreme conclusions about self, the environment, or community, and unnuanced reflections on the meaning of events. We linked this code to *person* as a subcode because it helps capture the personal reaction to the experience via absolute magnitudes.

Origin: This code was inductive and first applied to the Nina transcript but was discerned even earlier.

Importance: *Absolutist language* is important to the study because it calls attention to extreme situations and strong conclusions and shows how some people might see the hurricane and its outcomes in polar magnitudes.

Example:

I had never seen anything like that in my twenty years of living. (Nina)

It was completely calm outside, almost unbelievable. (Yvonne)

Many people, just like me, were left with absolutely nothing. (Mina)

Counterexample: Reference to something as "overwhelming" is not considered "absolutist" unless a word such as "completely" precedes it.

Reflection: *Absolutist language* reveals how the hurricane not only creates perceptions of extreme environmental circumstances, it creates extreme reflections regarding social forces and community as well. For example:

I noticed people cooking outside and I met many of my neighbors that I never really had gotten to know. (Mina)

I lived in Burlington, North Carolina my entire life and I have never experienced a disaster that I could come close to comparing with this. (Gina)

Beliefs

Description: The code *beliefs* refers to any language referring to an assumption or thought that a participant holds as true. It is an evident state of mind in which someone expresses credence or a tentative or confident statement.

Origin: This inductive code was first applied to Gary's transcript.

Examples:

I knew that we lived close to a creek, but not close enough for the water to get in the house. That is the same mistake so many people made. The idea that the water won't come any higher. It never has before. (Gary)

It was a horrible disaster, but the compassion shown was unbelievable. It makes me feel good to think that there are still some people who are

willing to do anything they can to help their neighbors during a hard time. (Sheila)

Counterexample: The following is a reflection on the hurricane, but it does not illustrate an identifiable belief, only a perception that the small-town atmosphere was wonderful.

In the middle of all the destruction was a wonderful human spirit that left no one feeling hopeless. That small-town atmosphere could not have been more evident than during those couple of days. (Gary)

Religion

Definition: *Religion* refers not only to places of worship, such as churches and synagogues, but to references to praying, faith, being blessed, believing in a higher power as well as specific references to God, Satan, Jesus, or other religious personae. We make a distinction between the code *religion* and the code *beliefs*. *Religion* is applied to beliefs and actions regarding a higher power or faith tradition (e.g., "God never gives you more than you can handle"), whereas *beliefs* is applied to more general beliefs (e.g., "It takes a disaster to bring people together.")

Origin: This code was inductive and first applied to Nina's transcript.

Importance: This code matters to the study because it points to religious beliefs, practices, and spiritual or religious behaviors that clarify how survivors use a religious meaning system to make sense of the flood. We can see whether religious sentiments show up before, during, and after the flooding. References to religion also reveal how people make religious meaning of the disaster.

Example:

Through many prayers, we attended the most emotional and uplifting game of my life. (Nina)

Reflection: During analysis, we will be interested in seeing whether and how those who reference praying also talk about helping others (i.e., the connection between *religion* and *action*). We will also be interested in seeing what, if anything, seems to activate praying. Also, does witnessing loss activate mentions of praying and religious meaning making?

Actions

Definition: *Actions* refers to behaviors and processes evident in the data, such as helping others, traveling, making phone calls, and talking to neighbors. We also use it as an organizing or "umbrella" code to organize subcodes referencing more specific actions, such as *assessing environment*, *decision making*, and *experiencing loss*.

Origin: This was an inductive code that we first applied in Orion's transcript.

Importance: As we began reading data, we noticed a large range of actions relevant to the larger notion of experience. Participants act and are acted upon; these actions and behaviors will help us better understand differences in how people respond to crisis and to people close to them.

Example:

Our generator blew up today and my family hopped into the car and drove all the way to Selma to trade it in for a new one. We had to drive all the way to Selma, an hour drive, because our Lowe's ad Tractor store in Rocky Mount were both approximately fifteen feet under water. Late this afternoon the power came on but only to be gone the next day when we woke up. I was supposed to work from 3 to 9 at Eckerd today but I called

in and told my boss that there was no way for me to get there. The roads were still blocked. (Orion)

Assessing environment

Definition: The code *assessing environment* refers to descriptions of hurricane damage and the writers' assessment of their physical environment. This includes descriptions of the devastating and destructive effect the hurricane had on the environmental setting, infrastructure, and persons. Though this may not seem like a traditional behavior or action—such as escaping a flooding house—we placed it under *actions* because assessing the environment requires using one's senses even if it is largely passive.

Origin: We first noticed this inductive code in Yvonne's transcript.

Importance: This code will allow us to keep track of the participant as an eyewitness of the flooding. That is, we will be able to see what kind of lens the participant uses as they describe their surroundings.

Examples:

Rivers have risen an unbelievable tremendous amount since yesterday. Apparently Floyd dumped an enormous amount of rain on eastern NC and showed no mercy. Houses and roads are flooded from the rain yesterday and now rivers are rising. A group of friends and I had made plans to head to Salem, VA for the weekend. What should have been a four-hour trip turned out to be seven hours due to closed roads. Around 11:30 p.m. we crashed into our motel rooms and awaited the morning. (Lana)

People are still taking a ticket and standing in line to get in the store for a few minutes.

There were trees on houses, on cars, on boats that were brought in and trees in the road. The street in front of campus was a flowing river I could litualy [sic] take out a boat. (Robin)

Decision making

Definition: The code *decision making* refers to language suggesting that someone has made or is making a choice regarding how to manage the hurricane, their emotions, or others' safety.

Origin: This code was a deductive, provisional code (Saldaña, 2016), a topic we anticipated finding in the data.

Importance: This topic is closely linked to how participants experience a hurricane, our central research question. How they weigh options in making a decision is a critical component of their larger experience.

Example:

> I live in an apartment here in Greenville. With Hurricane Floyd approaching, I decided to return to my mother's house, along with my girlfriend, to ride out the storm. She lives about 45 minutes away in Deep Run, outside of Kinston. (Michael)

Anti-example: General decision making should not be coded.

> Last semester I decided to declare my major, elementary education. (Gina)

Reflection: *Decision making* will also allow us insights into participants' awareness, use of, and satisfaction with the support available, including possibly local and federal government assistance as well as informal support.

Experiencing loss

Definition: *Experiencing loss* is applied to any text segment including the word "loss" or "lost" as well as similar terms. The loss can be specific and concrete (e.g., food) as well as general or abstract, such as losing one's self confidence. Loss can refer to loss of neighbors (who were displaced), community, and stability as well as tangible property loss.

Origin: *Experiencing l*oss was a deductive code, though we had not considered issues such as food loss, which can be a financial as well as emotional challenge. As a deductive code, we titled this topic *loss* but later decided that *experiencing loss* better captures the meaning of this code and how it captures what is happening to the participants.

Examples:

> We started loading as much stuff as we could because she knew that she was going to probably lose the first floor. We stored stuff on the second level thinking that the water wouldn't get that high. Little did we know it would. (Diana)

> Luckily, the only thing we lost was a lot of food due to the power outage. I will take a loss of food any day. (Nina)

> The one thing I did lose was the small amount of structure that I had in my life. Since I moved to Greenville in 1997, I have tried to gain some sort of stability. (Gina)

Counterexample: Even if a house or apartment is described as flooded, it was not coded as *loss* unless loss was referenced more explicitly.

> When we left Tar River there was close to three feet of water in her

apartment. I felt so bad for her but I was glad that I had a place for her to stay. She brought her stuff inside and I didn't know how long she would be with us. (Diana)

Reflection: It is interesting to note that loss can be anticipatory as well as actual. Even as something anticipated, it can generate anxiety. Hence, we might say that loss has the properties of tangibility (from tangible to abstract) and anticipation (from anticipatory to surprising). Loss is also frequently contextualized. That is, it is recognized in comparison to what others have lost or to what could have been lost.

Helping others

Definition: This code refers to talking about help as well as providing or receiving help because of hardship from the flood. This code should be applied even if it is not the participant who is offering the help or receiving assistance.

Origin: This code was inductive and first noticed in Sheila's transcript.

Importance: This topic will provide valuable insight into prosocial behavior and when helping others was and was not evident.

Examples:

Those people needed my help, not my tears and sympathy. (Sheila)

People seemed to come out of the woodwork to help those in need. (Sheila)

Counterexample: This code will not apply to references to help that were unrelated to the hurricane (e.g., helping someone with their homework).

Reflection: This code is especially rich because it will help us understand not only

Learn to Build a Codebook for a Generic Qualitative Study

giving and receiving assistance but how people decide that some situations are worthy of their support while others are not. In other words, this topic is an entry point into related topics, such as how events are construed as emergencies or not, how people develop feelings of responsibility, and if/how they decide they have the necessary skills to help.

Socializing

Definition: This code will be applied to incidents of attending hurricane parties, talking to family or friends, and spending time with neighbors in general conversation. We make a distinction between *communication* and *socializing*. *Communication* refers to efforts to communicate regarding the hurricane or one's whereabouts, whereas *socializing* simply refers to spending time with others.

Origin: This code was inductive and first noticed in Rena's transcript.

Importance: It will be helpful to draw attention between socializing and other forms of interaction with friends, family, and neighbors. Noticing when people are isolated versus when they are communicating with others will help us analyze differences between these two modes. How are experiences of participants spending time with others different from those who did not?

Example:

> When we got there we had a small thrown together birthday party for me, which felt wierd (sic) being that I didn't even feel like it was my birthday. (Rena)

Counterexample: Text segments regarding rescuing or receiving aid will not be coded as *socializing*. Rather, they will be coded as *helping others*.

Trying to live normally

Definition: Continuing to live normally or not realizing until having to face reality; trying to ignore the situation with various levels of success. Excerpts depicting participants who try to "look away" from the damage would fall into this code.

Origin: This code was an interpretive code based on our review of both quotations as well as holistic narratives. It captures a perspective whereby individuals do their best to not let the flooding completely derail their plans. Individuals who proceed with their birthday plans or drive to another state and avoid the media would be examples of trying to live a normal life.

Importance: This code is critical to understanding how people try to maintain life as they know it as the world around them is changing dramatically.

Examples:

> We ended up going to a big hurricane party and were having a blast for some time. We were oblivious to what was going on outside. Then the power went out which brought us back to reality. (Nila)

> All she thought was to clean the carpets, wipe down the walls and begin to get life back to normal. That wasn't the case. (Gary)

Reflection: This code has the potential to become a larger theme connecting other codes. That is, code combinations can help us further assess evocative concepts such as *trying to live normally*. By looking at the overlap of codes such as *socializing* and *decision making*, we might see that these topics in concert suggest *trying to live normally*. A codebook is not the place where we would develop a theme; that analytical effort would instead happen in a separate memo.

Emotions

Definition: The code refers to emotional reactions of the narrator, both positive

and negative, evaluative statements or descriptions. As a general topic, *emotions* function as an organizing code under which other codes lie, such as *compassion*, *lack of control/agency*, *gratitude*, *luck*, and *negative emotions*.

Origin: This provisional code was part of our deductive code list.

Importance: As a broad code, *emotions* will allow us to understand the range of emotions evident in the data. Ultimately, we will query data based on the co-occurrence of *emotions* and other codes, such as *actions* in order to generate multiple-concept codes or themes.

Example:

> Television reporters are saying that a flood like this happens once every five hundred years and this is the worst one of the century. It is an extremely sad situation.

Compassion

Definition: This code refers to compassion, empathy, and sympathy evident in the narrator or others.

Origin: This was an inductive code noticed first in Robin's transcript. It can be applied as an interpretive code; see example below.

Importance: This code, like the *helping others* code, will help us assess feelings of goodwill and how this sentiment is expressed generally to the community or for particular individuals.

Example:

> The look on their face said it all. I felt their pain. (Robin)

Gratitude

Definition: This code refers to any indications of being grateful regarding assistance, kindness, outcomes, gifts, favors, or other types of generosity. Both implicit and explicit mentions will be coded.

Origin: This was an inductive code that we first noticed in Mary's transcript.

Importance: This code will help us make sense of positive emotions in the data.

Example:

> I just thank God that he protected me and my family and I've learned to appreciate what I have. (Mary)

Lack of control/agency

Definition: Not being able to control the forces of nature; reacting to the hurricane but feeling that one's actions are inadequate; being at the mercy of the hurricane.

Origin: This was an inductive code that we started working with in Orion's transcript.

Importance: This code matters to the study because it will help us assess whether people are aware of a loss of control and what they do as a result. Control is an instrumental part of behavior; hence, this code will allow us to better dimensionalize action in the data and compare actions generated from feeling in control to those generated from lack of control.

Examples:

> During the trip we saw merely the roofs of homes and their inhabitants leaving their homes by boat. I felt helpless. (Orion)

> I could not help them because the roads were beginning to flood.

Luck

Definition: The code luck is applied to text segments referencing being lucky, fortunate, or unfortunate. Mentions of feeling "blessed" will be included as well.

Origin: This code was an inductive code first applied to Mitchell's transcript.

Examples:

> I was so incredibly thankful that my family had faired so well, but I was also filled with a feeling of guilt that so many others had not been as lucky as I. (Mitchell)

> I was very lucky in the fact that my house did not receive any damage; my mother and I lost no shingles or even had any trees to fall. We consider ourselves extremely fortunate.

> Seeing all of these people who lost everything in such good spirits made me realize how truly blessed I was.

Reflection: It is interesting to note that people who refer to luck reveal related attitudes. Some participants express luck and talk in detail about what others lost while others express luck and talk mostly about themselves. Analyzing *luck* further will help us assess the range of self-focused to other-focused perspectives.

Negative emotions

Definition: This code refers to negative feelings such as guilt, fear, and overwhelm.

Origin: We decided to make a code specifically for negative emotions in order to distinguish them from emotions such as compassion and gratitude.

Importance: This code will help us make sense of the range of negative emotions and whether certain negative emotions are more prominent than others.

Learn to Build a Codebook for a Generic Qualitative Study

Example:

> I was so incredibly thankful that my family had faired so well, but I was
> also filled with a feeling of guilt that so many others had not been as lucky
> as I. (Mitchell)

Communication

Definition: This code is applied broadly to modes of communication (such as calling) and attempts to communicate regarding the hurricane or one's emotional or physical state. Communication acts as an organizing or "umbrella" code to help us tie together the subcodes *media* and *messaging*. It was also applied to descriptions of attempts to communicate, such as calling a friend or relative.

Origin: This code was inductive and one that we decided to use after reading accounts of participants using various modes of communication, such as phones (e.g., Diana), to spread information about the hurricane and their circumstances.

Importance: This code will allow us to compare different modes of communication (i.e., phone, e-mail, conversation) as survivors attempt to make sense of the environmental devastation and personal risks involved in taking care of friends and family—or in reaching out for support.

Example:

> Next we went to Wildwood Apartments on Beech St. to see my good
> friends Sara & Teresa' their phone had been out all night. (Kat)

Reflection: We are interested in the function of communication before, during, and after the hurricane. What happens when attempts to communicate fail? Are there differences in communicating by phone versus direct interaction?

Media

Definition: This code refers to references to any news media—radio, television, newspaper, and online sources.

Origin: *Media* was a deductive code.

Importance: This code is important because it not only references whether people mention the media but what role the media plays in their decision making, outcomes, and relationship with others. In other words, it is an important contextual code that gives shape to a larger messaging context.

Examples:

> We watched the news and realized that this one wasn't going to be like the last hurricane. (Mina)

> I realized how much we rely on the media for our information. Depending on what they predicted, I would choose to evacuate or not. (Sonya)

Reflection: The media can have great influence in how individuals make sense of their experience. Just as conversations with friends and family can be pivotal, a turning point can also happen by a news announcement that brings home the severity of the storm.

Messaging

Definition: The code *messaging* applies to language suggesting information about the hurricane that is accepted without question. This is different from the code *media*, which we will use to track specific topics such as television, radio, and online media sources. *Messaging*, in contrast, is a broader topic. In some cases, the informational source is vague. In other cases, it is mentioned as a specific person, such as a parent.

Origin: This code was first applied in Yvonne's transcript.

Learn to Build a Codebook for a Generic Qualitative Study

Importance: Messaging is important for the study because it influences how individuals make decisions and how they make sense of their experience. It will also allow us to see how a range of message types were received.

Examples:

There is word floating around of another hurricane headed our way. (Yvonne)

The news said this morning that the hurricane is headed for us. (Lana)

Reflection: There is an implied hierarchy of messaging suggesting a structure of authority. The official word from the media, rumor from friends, "word from campus," parental requests, and information from romantic partners and roommates build a hierarchy that we can further analyze. Each of these authorities provide information that participants use to make decisions and act most prudently.

Environment

Community

Definition: The larger community, including persons who have no social relationship with the narrator; "faceless," "distant," and impersonal people.

Importance: This code will allow us to compare how the participants characterize close friends and family versus people they barely know or the community or neighborhood more generally.

Example:

I have never seen so many frantic people before.

Property

Definition: Property refers to houses, cars, furniture, as well as smaller objects. It can also refer to general references to property, such as "I can't believe how many homes were damaged." An initial code titled *valuable objects* was merged with *property* because the distinction between these two topics was not important for our research question. *Valuable objects* referred to specific objects that participants dwelled upon, such as a jewelry box.

Origin: This provisional code was part of our deductive list. We expected participants to discuss the physical property that was lost or damaged.

Importance: It is important to consider not just references to property but people's relationship to property. This will allow us to analyze related topics such as materialism and whether and how what people own is related to their self-image.

Examples:

> They tried to drive through some water on the road and the water started rushing into the car. They left the car and waded waist high the rest of the way. (Mina)

> Inside you could see that the water had gotten about 15 inches high by the line of leaves, grass, and whatever else was flowing on top of that water. At first my mother was pleased, she had saved most of her furniture, all of her pictures, and her important documents. (Gary)

Counterexample: This code should not be applied to references to food or supplies.

> My mother also left work to search for those items that were nonexistent in almost every grocery story around.

Timescale

Definition: This code is broadly applied to any words suggesting time—today, yesterday, tomorrow, before, since, during. It also refers to time markers such as mentioning a particular day or span of days. It was linked to environment because we decided that *timescale* is a kind of temporal environment and contributes to how one experiences settings and spaces.

Example:

Today, we were informed that classes were cancelled because of the closeness of Hurricane Floyd. (Nina)

What should have been a four-hour trip turned out to be seven hours due to closed roads. (Yvonne)

Counterexample: Even though "since" and "then" are words suggesting time, if they occur in sentences having nothing to do with time in relation to the hurricane, then the code would not be applied.

Reflection: In analysis, we would like to analyze the textual segments for this code for references to delayed time—e.g., how time seems to be standing still. We might also assess differences in timescales across data. Some individuals give accounts based on moment by moment incidents, while other people use a different timescale. Understanding what kind of participants use a particular timescale can be elucidating and may later suggest a theme.

People

Definition: This is a broad, placeholder code to organize subcodes beneath it. No data will be coded as *people*. They will be coded with the specific subcodes.

Family

Definition: Family members of the narrator.

Importance: Coding for family members will allow us to assess how participants make sense of the hurricane and flooding in relation to their parents.

Example:

> Both of my parents had to work and would not be able to go to stop and gather necessitates for our family's survival.

Counterexample: We will not include references to people who are described as being "like family." Nor will we consider general references to families, such as the following:

> When I see the news stories of families that have lost everything, I wonder how they will ever recover because the majority of them were barely scraping by before the flooding.

Reflection: I wonder whether the hurricane heightens tension between parents and their children or whether tension dissipates in light of the severity of the disaster. Reviewing quotations coded to this code will help us address this question.

Assessing Codebook

These codes are based on a generic research approach rather than a more specific tradition such as grounded theory, narrative analysis, or discourse analysis, each of which would have its own coding focus. Grounded theory would have numerous initial codes. Narrative analysis would have codes specifically focused on components of stories, such as *complicating action* and *resolution* as well as *settings, protagonists*, and *antagonists*. In a generic qualitative study, we should ask ourselves whether the codes give us room to move forward with

inquiry. After having constructed a codebook and having applied codes to data, the next step would be to run reports of each code to look for patterns within each topic and then to assess combinations of codes in constructing themes. Put another way, our codes should give us direction in asking further questions of the data as well as in assessing multiple concepts and well as single concepts. These analytical steps are beyond the scope of this dataset, but related datasets provide examples of analysis.

Review

Qualitative analysis incorporates a variety of tools including memo writing, diagrams, and coding. Applying codes to textual segments is a way of focusing one's attention on condensed topics most relevant to the study's research questions (Saldaña, 2016). Coding is a way to keep our eyes on what matters as we systematically review text segments and increasingly mine them for meaning. In confirmatory research, an *a priori* approach is adopted whereby codes and how they are operationalized have been established perhaps even before data collection. In contrast, in exploratory work, the topics of discovery are continually refined as inductive codes are developed and applied to more and more data. In some cases, a code might not be discernible across a sufficient number of transcripts for it to be part of the final analysis. In other cases, the code might need to be renamed or refined.

After completing this article, you should be able to:

- Understand how coding is used in qualitative analysis.
- Define codes.
- Understand how codebooks are developed and used.

Reflective Questions

Learn to Build a Codebook for a Generic Qualitative Study

1. Is your study driven by a conceptual framework or more by the data itself?

2. Can you identify 5–10 *a priori* codes for your study—topics that you anticipate finding in the data?

3. Are your codes primarily descriptive (tangible) or will some of them also be interpretive (abstract or theoretical)?

References

Maietta, R., **Hamilton, A.**, **Swartout, K.**, & **Petruzzelli, J.** (2018). ResearchTalk's Qualitative Data Analysis Camp. Carrboro, NC. Retrieved from http://www.researchtalk.com/qualitative-data-analysis-camp-april-2018/

Saldaña, J. (2016). *The coding manual for qualitative researchers* (3rd ed.). Thousand Oaks, CA: SAGE.

Strauss, A., & **Corbin, J.** (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Thousand Oaks, CA: SAGE.