



## Thematic analysis of qualitative data: AMEE Guide No. 131

Michelle E. Kiger & Lara Varpio

To cite this article: Michelle E. Kiger & Lara Varpio (2020): Thematic analysis of qualitative data: AMEE Guide No. 131, Medical Teacher

To link to this article: <https://doi.org/10.1080/0142159X.2020.1755030>



Published online: 01 May 2020.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

## Thematic analysis of qualitative data: AMEE Guide No. 131

Michelle E. Kiger<sup>a,b</sup>  and Lara Varpio<sup>a,b</sup> 

<sup>a</sup>Wright-Patterson Medical Center, Dayton, OH, USA; <sup>b</sup>Uniformed Services University of the Health Sciences, Bethesda, MD, USA

### ABSTRACT

Thematic analysis is a widely used, yet often misunderstood, method of qualitative data analysis. It is a useful and accessible tool for qualitative researchers, but confusion regarding the method's philosophical underpinnings and imprecision in how it has been described have complicated its use and acceptance among researchers. In this Guide, we outline what thematic analysis is, positioning it in relation to other methods of qualitative analysis, and describe when it is appropriate to use the method under a variety of epistemological frameworks. We also provide a detailed definition of a *theme*, as this term is often misapplied. Next, we describe the most commonly used six-step framework for conducting thematic analysis, illustrating each step using examples from our own research. Finally, we discuss advantages and disadvantages of this method and alert researchers to pitfalls to avoid when using thematic analysis. We aim to highlight thematic analysis as a powerful and flexible method of qualitative analysis and to empower researchers at all levels of experience to conduct thematic analysis in rigorous and thoughtful way.

### KEYWORDS

Thematic analysis; qualitative research methods; qualitative analysis

### Introduction

Data analysis has been described as 'the most complex and mysterious of all of the phases of a qualitative project, and the one that receives the least thoughtful discussion in the literature' (Thorne 2000). Many qualitative research papers lack explicit description of the methods informing data analysis, or, when included, the terms used to describe data analytic methods are often used imprecisely or are mislabeled entirely (Sandelowski and Barroso 2003; Sandelowski 2010). Further complicating matters, certain terms describing qualitative data analysis have either carried a wide range of definitions or lacked clear definitions. This imprecision leads to a lack of transparency, making it difficult for readers to understand how data analysis was performed and, consequently, how to interpret findings (Nowell et al. 2017). It also contributes to perceptions that qualitative research is less rigorous than quantitative research (Clarke and Braun 2013).

Unfortunately, this lack of clear terminology plagues a qualitative data analysis method that is among those most frequently used in health professions education (HPE) research: *thematic analysis*. *Thematic analysis* is a term that has been variably defined (Merton 1975; Aronson 1995; Boyatzis 1998; Attride-Stirling 2001; Braun and Clarke 2006; Joffe 2011), and that has even been discounted as unsophisticated or inferior to other qualitative methods (Braun and Clarke 2006, 2014). Many researchers who use thematic analysis fail to provide sufficient descriptions of the analysis process followed and of the theories or epistemological assumptions undergirding the analyses (Attride-Stirling 2001; Braun and Clarke 2006). Additionally, many studies that have employed thematic analysis have not explicitly labeled it as such in their manuscripts; instead, these reports simply state that 'qualitative data were examined for recurring themes', without offering

### Practice points

- Thematic analysis is a powerful yet flexible method for analyzing qualitative data that can be used within a variety of paradigmatic or epistemological orientations.
- Thematic analysis is an appropriate method of analysis for seeking to understand experiences, thoughts, or behaviors across a data set.
- Themes are actively constructed patterns (or meanings) derived from a data set that answer a research question, as opposed to mere summaries or categorizations of codes. Themes can be generated inductively or deductively.
- The most widely-accepted framework for conducting thematic analysis involves a six-step process: familiarizing yourself with the data, generating initial codes, searching for themes, reviewing themes, defining and naming themes, and producing the report.
- Given the flexibility of thematic analysis, researchers using this method must clearly outline their paradigmatic orientations and assumptions to ensure the trustworthiness of their findings and interpretations.

further explanation (Braun and Clarke 2006). Clearly there is considerable confusion amongst researchers about what thematic analysis means, when to use it, and how to use it.

Thematic analysis is a practical data analysis approach for qualitative researchers; clarifying how to use it

appropriately and effectively can help HPE researchers recognize its utility, versatility, and power. In this Guide, we aim to support the achievement of these goals. First, we define thematic analysis, focusing on the flexibility that it offers researchers. We explore how it can be applied across a range of theoretical and epistemological frameworks. We also suggest when thematic analysis can be harnessed in qualitative data analysis. Next, we focus on some key concepts underpinning thematic analysis. Specifically, we discuss the definition of a theme, including different types of themes (i.e. semantic versus latent), and how inductive or deductive processes can be employed to develop themes. We then describe a stepwise approach for conducting thematic analysis, following the six-step framework of Braun and Clarke (2006) and providing a worked example from our own research data to illustrate each step. We conclude with a discussion of the advantages and disadvantages of using thematic analysis, and a description of pitfalls to avoid.

### **What is thematic analysis?**

Thematic analysis is a method for analyzing qualitative data that entails searching across a data set to identify, analyze, and report repeated patterns (Braun and Clarke 2006). It is a method for describing data, but it also involves interpretation in the processes of selecting codes and constructing themes. A distinguishing feature of thematic analysis is its flexibility to be used within a wide range of theoretical and epistemological frameworks, and to be applied to a wide range of study questions, designs, and sample sizes. While some scholars have described thematic analysis as falling within the realm of ethnography (Aronson 1995) or as particularly suited to phenomenology (Joffe 2011), Braun and Clarke (2006) argue that thematic analysis can stand alone as an analytic method *and* be seen as foundational for other qualitative research methods. Indeed, the principles of thematic analysis of how to code data, to search for and refine themes, and to report findings are applicable to several other qualitative methods such as grounded theory (Watling and Lingard 2012) and discourse analysis (Taylor et al. 2012). Because of this flexibility, Braun and Clarke (2006) refer to thematic analysis as a method, as opposed to a more tightly prescribed methodology.

Thematic analysis is not bound to a particular paradigmatic orientation; instead, it can be used within post-positivist, constructivist, or critical realist research approaches (Braun and Clarke 2006). Using thematic analysis in different research paradigms entails harnessing this method to distinct purposes and outputs. Post-positivists can use thematic analysis to focus on individuals' meanings and experiences to gain insights into the external reality, thereby supporting the development of conjectural knowledge about reality. In many interpretivist orientations (e.g. constructivism), thematic analysis can emphasize the social, cultural, and structural contexts that influence individual experiences, enabling the development of knowledge that is constructed through interactions between the researcher and the research participants, revealing the meanings that are socially constructed (Braun and Clarke 2006). Joffe (2011) suggests that thematic analysis is particularly suited

to constructivism because, through the process of analyzing a wide range of data, it can illustrate how a certain social construct develops. In these ways, constructivist thematic analyses will search for more latent, deeper themes within the data. Finally, critical realism acknowledges experiences and perceptions grounded in a material reality but seeks to investigate social meanings and implications behind the topic of interest (Joffe 2011; Clarke and Braun 2017). Within a critical realist framework, thematic analysis can allow researchers to study the power relations informing reality and to engage in emancipatory investigations that value the voices of oppressed populations.

Among those who have described thematic analysis as a post-positivist method (Aronson 1995; Boyatzis 1998), Boyatzis (1998) forwards thematic analysis as a method that can bridge the chasm between the post-positivist pursuit of understanding a reliable, objective, fact-based reality, and the more interpretive aims of many social science researchers. Boyatzis posits that 'thematic analysis allows the interpretive social scientist's social construction of meaning to be articulated or packaged in such a way, with reliability as consistency of judgment, that description of social "facts" or observations seems to emerge' (p. xiii). He suggests that the interplay between post-positivist and interpretivist paradigms within thematic analysis can produce a symbiosis in which interpretive findings can generate new hypotheses to be tested using post-positivist methods, and post-positivist hypothesis testing can in turn suggest new themes for exploration from an interpretive lens.

### **When to use thematic analysis**

Thanks in large part to those who have clearly laid out its analytical processes (Braun and Clarke, 2006, 2012; Clarke and Braun 2017), researchers have suggested that thematic analysis is a good first analytic method for novice qualitative researchers to master (Braun and Clarke 2006, 2012; Clarke and Braun 2017; Nowell et al. 2017). However, as with any research or analytic method, we would argue that the choice to use thematic analysis should be based on the goals of the research itself, more than a desire to select an easy-to-follow method of analysis. Thematic analysis is an appropriate and powerful method to use when seeking to understand a set of experiences, thoughts, or behaviors across a data set (Braun and Clarke 2012). Since it is designed to search for common or shared meanings, it is less suited for examining unique meanings or experiences from a single person or data item. Finally, because of its relevance to other methods of qualitative research, the steps of thematic analysis echo those of grounded theory, ethnography, and other qualitative methodologies that also rely on coding and searching data sets for themes as part of their processes.

Situating thematic analysis in relationship to other qualitative analysis methods can help us understand the method's scope and purpose. The framework offered by Sandelowski and Barroso (2003) is a useful lens through which to compare and contrast such methods. Sandelowski and Barroso (2003) contend that qualitative analysis methods fall along a continuum defined by the degree to which data is transformed during analysis. This continuum is

grounded at one pole with purely descriptive analyses in which the data is not significantly transformed. Analysis methods at this far end include, for example, *topical surveys* which Sandelowski and Barroso (2003) argue should not be classified as true qualitative research because they focus on reporting lists or inventories of topics raised by interview or focus group participants, often as frequencies or percentages, but make little or no effort to purposefully sample participants or interpret findings. At the other end of the continuum are highly interpretive analyses in which there is considerable transformation of the data. Located at this pole are methods, such as phenomenology, which involve transformation and deep interpretation of data. Interpretative phenomenological analysis looks in detail at how individual experiences and the meanings that people attach to them can inform a question of interest (Smith and Osborn 2003).

We suggest that, while thematic analysis can be used across the continuum, it most naturally lands near the center between the two poles. Through thematic analysis, the research constructs themes to reframe, reinterpret, and/or connect elements of the data. Thus, themes are not merely organizational tools used to classify and label data. While processes of thematic analysis will have the researcher developing organizational and classification labels to describe the data, thematic analysis goes further into the interpretation and data transformation processes. But if thematic analysis does not belong at the purely descriptive pole of the analysis continuum, it also does not belong at the highly interpretive pole. Thematic analysis is generally not used to engage in data interpretation and transformation to the point of developing theory, the central goal of grounded theory (Glaser and Strauss 1967). Instead, thematic analysis lands most naturally between the poles—engaging in more than description and categorization, but not extending so far as to develop theory.

### What is a theme?

Before delving into the specific steps of thematic analysis, it is important to define what the term *theme* means in this analysis method. A theme is a ‘patterned response or meaning’ (Braun and Clarke 2006, p. 82) derived from the data that informs the research question. Viewed in opposition to a category—which provides description and organization to the ‘manifest content’ of a data set—a theme is a more abstract entity that involves a greater degree of interpretation and integration of data (Nowell et al. 2017). When engaging in thematic analysis, researchers can identify themes irrespective of the number of times a particular idea or item related to that theme appears in a data set. Furthermore, the importance or centrality of a theme is not necessarily reflective of the frequency of its appearance within the data (Braun and Clarke 2006; Nowell et al. 2017). Themes can be classified as either *semantic* (also often labeled as *manifest*), which address more explicit or surface meanings of data items, or *latent*, which reflect deeper, more underlying meanings, assumptions, or ideologies (Boyatzis 1998; Braun and Clarke 2006). The researcher has great flexibility in which themes to identify, but he or she should strive to identify themes that provide important

insights that address the research question (Braun and Clarke 2006).

Researchers can employ an *inductive* or *deductive* approach to theme identification (Braun and Clarke 2006, 2012). An *inductive* approach, as used in grounded theory, derives themes from the researcher’s data (Varpio et al. 2019). Since these themes are data driven, they might not mirror the exact questions asked of participants (e.g. if participants veered off topic), and they are not necessarily reflective of the researcher’s own interests or beliefs on the subject (Braun and Clarke 2006). Conversely, *deductive* approaches use a pre-existing theory, framework, or other researcher-driven focus to identify themes of interest (Braun and Clarke 2012; Varpio et al. 2019). Therefore, an inductive approach tends to provide a broader, more expansive analysis of the entire body of data, whereas a deductive approach is useful for honing in on a particular aspect of the data or a specific finding that could be best illuminated or understood in the context of a pre-existing theory or frame (Braun and Clarke 2006). While either method is acceptable, specifying the approach used is important to allow readers to properly interpret and contextualize findings.

### How to engage in thematic analysis

Several researchers have published descriptions and guides of how to conduct different versions of thematic analysis (Aronson 1995; Boyatzis 1998; Attride-Stirling 2001; Joffe 2011). In this guide, we will focus on the method as outlined by Braun and Clarke (2006) as it has become the most widely adopted method of thematic analysis within the qualitative literature (Clarke and Braun 2017). Their method of analysis consists of six steps. It is important to note that Clarke and Braun’s thematic analysis is designed to be a recursive, rather than linear, process in which subsequent steps may prompt the researcher to circle back to earlier steps in light of new data or newly emerging themes that merit further investigation.

To illustrate these steps, we offer an example using original data from a study we performed examining the experience of patient ownership in continuity clinics within a pediatric residency program (see Box 1 for illustrations of each step’s transformation of the data). Based on a scoping literature review, we (MK, LV, and others) have proposed a definition of patient ownership as ‘the commitment that a medical provider—both individually and as part of a team of healthcare professionals—feels and displays in relation to the provision and coordination of care for his or her patients’ (Kiger et al. 2019). However, recognizing that personal experiences of patient ownership will inevitably be shaped by subjective experiences and context, we conducted individual interviews of residents, attending physicians, and patient families to understand definitions, experiences, and expectations of patient ownership from these different perspectives. In this example, we employ an *inductive* approach to thematic analysis, and work within a *constructivist* epistemology.

#### Step 1: Familiarizing yourself with the data

The first step in thematic analysis’s process is becoming familiar with the entire data set, which entails *repeated* and

**Box 1** Worked example of thematic analysis from patient ownership project.

## Step 1: Familiarizing Yourself with the Data

Below is an excerpt from a resident transcript that we will use to illustrate the steps of thematic analysis.

'So, I think there's a couple of families that I've taken ownership on, and part of it is, I think – like, I just like them so much. They're such nice people. Not that like, I don't take ownership of the patient families that aren't nice, but the ones that I like, truly felt like, "Oh, I want you to come see me and only me" are the ones that you really like, click on a personal level with. And whether it's like, not even just like, "Oh, the mom and I could be friends outside of here or the dad and I could be friends outside of here" but there's just something about the family dynamic and the kid that you watch, and you get to watch grow. And I think, also, like, the younger they are, the more of that that there is, "cause you get to like, see all of that, and you're like, "Oh, come back in two months and see me so I can see what you're doing then.'" Resident 1

## Step 2: Generating Initial Codes

*Codes identified in excerpt:*

**Physician feelings toward patients:** 'So, I think there's a couple of families that I've taken ownership on, and part of it is, I think – like, I just like them so much. They're such nice people. Not that like, I don't take ownership of the patient families that aren't nice, but the ones that I like, truly felt like, "Oh, I want you to come see me and only me" are the ones that you really like, click on a personal level with. And whether it's like, not even just like, "Oh, the mom and I could be friends outside of here or the dad and I could be friends outside of here" but there's just something about the family dynamic and the kid that you watch, and you get to watch grow'.

**Intrinsic sense of responsibility:** 'Not that like, I don't take ownership of the patient families that aren't nice, but the ones that I like, truly felt like, "Oh, I want you to come see me and only me" are the ones that you really like, click on a personal level with'.

**Continuity of care:** 'And I think, also, like, the younger they are, the more of that that there is, "cause you get to like, see all of that, and you're like, "Oh, come back in two months and see me so I can see what you're doing then.'"'

*Coding manual examples:*

**Physician feelings toward patients:** the feelings, either positive or negative, that a physician has toward his or her patients

**Intrinsic sense of responsibility:** the sense of responsibility that a physician feels toward his/her patient, as opposed to the requirements imposed upon them by attending physicians, systems, or clinic-wide expectations or policies

**Continuity of care:** from the physician's perspective, seeing the same patient longitudinally; from the patient's perspective, seeing the same physician longitudinally

## Step 3: Searching for Themes

Noticing that continuity of care was frequently mentioned as an important facilitator of patient ownership, we developed one theme to address the connection between these two concepts. Continuity of care facilitated relationship building, but participants also provided examples of how having continuity of care did not always translate to better ownership, and, conversely, of how some physicians had excellent patient ownership without having continuity of care. Therefore, it appeared that continuity of care was valuable but was neither necessary nor sufficient to guarantee patient ownership. The theme was designed to make connections between these concepts and meaningfully interpret the data. Notice that the theme was *not* a mere summary or categorization of codes, such as 'Effects of continuity of care' or 'Participant perspectives on continuity of care'.

Our initial theme: 'Continuity of care supports patient ownership but is not synonymous with patient ownership'.

The main theme was divided into two sub-themes: 'Importance of continuity of care' and 'Factors that mediate the relationship between continuity of care and patient ownership.' Our codes from above, 'Physician feelings toward patients' and 'Intrinsic sense of responsibility', were both considered mediating factors that affect the interplay between patient ownership and continuity of care and added to this sub-theme.

## Step 4: Reviewing Themes

In this step, additional codes from different portions of the transcripts were incorporated into the theme and its sub-themes.

'Educational value', 'Benefits to patient', and 'Benefits to patient' (which were not captured in the excerpt above but were taken from other transcripts) were added to 'Importance of continuity of care.' Beyond the patient-physician relationship, these three factors were seen as additional reasons why respondents valued continuity of care.

'Patient-level factors affecting ownership' (e.g. the complexity of a patient's medical history or presenting complaint), 'Resident career goals', and 'Lack of knowledge or training' were added as mediating factors. Physicians and patients saw continuity of care as less important for more straightforward medical problems or for patients who had less complex medical histories.

Residents were seen as more likely to want to take ownership of patients, especially those they did not see regularly, if they wanted to pursue a career in primary care, whereas they were less likely to do so when they felt they lacked specific medical knowledge or training (particularly for more complex patients).

## Step 5: Defining and Naming Themes

In this step, we finalized our themes and reworded our final definition to clarify the relationship we were proposing between continuity of care and patient ownership. We realized the main argument this theme was conveying was not simply that patient ownership and continuity of care are not synonymous, but rather that continuity of care is important but not sufficient to ensure patient ownership. Therefore, the final naming was: 'Continuity of care exerts a powerful influence on patient ownership, but it cannot guarantee patient ownership.'

We then again examined this theme in relation to the other four themes we had constructed:

1. Patient ownership involves a relational commitment between patient and physician that includes both affective and task-based components.
2. Patient families and physicians harbor idealized conceptions of patient ownership but acknowledge the constraints imposed by logistical and systems-based factors.
3. Patient families and physicians hold an expansive view of team-based ownership that includes physicians, support staff, and patient families.
4. Physicians recognized the importance of placing limits on the degree of ownership expected of residents.

(continued)

### Step 6: Producing the Report/Manuscript

We determined that the theme and its sub-themes did not have significant overlap with the other themes. It was distinct from them because it was focusing on a particular influencer of patient ownership, whereas the other themes were defining patient ownership (Theme #1 above), describing what it entails (#3), and delineating its limitations (#2 and 4).

Below is an excerpt from the manuscript in which this selection of the transcript was incorporated.

Note: this excerpt is only a portion of the section describing this theme.

'However, while continuity of care was identified as a facilitator for patient ownership, it was neither necessary nor sufficient to achieve patient ownership. Instead, several factors mediated the relationship between continuity of care and patient ownership.

'First, residents recognized that their personal feelings toward patients were key mediators of their sense of patient ownership. Even though residents acknowledged the need to provide good care and follow up to all patients regardless of their personal feelings, having a positive personal connection with families over time made it more likely they would actively work to take ownership of their care.

So, I think there's a couple of families that I've taken ownership on, and part of it is, I think – like, I just like them so much. They're such nice people. Not that like, I don't take ownership of the patient families that aren't nice, but the ones that I like, truly felt like, 'Oh, I want you to come see me and only me' are the ones that you really like, click on a personal level with. -Resident 1'

active reading through of the data (Braun and Clarke 2006). Depending on the study, the data set might include interviews, focus groups, recorded observations, field notes, journal entries, or other media such as photographs or videos (Thorne 2000; Nowell et al. 2017). While it can be tempting to begin coding data and searching for themes immediately, familiarizing oneself with the entirety of the data set first will provide a valuable orientation to the raw data and is foundational for all subsequent steps. For audio data that need to be transcribed, the process of transcription can be time-consuming but also serves as an excellent way to become familiar with the data. If researchers use voice-recognition software or hire transcription services to facilitate this step, checking the transcripts against original audio recordings for accuracy may be similarly useful.

### Step 2: Generating initial codes

As the first truly analytic step in the process, coding helps to organize data at a granular, specific level. After step 1's familiarization work, researchers can begin to take notes on potential data items of interest, questions, connections between data items, and other preliminary ideas. This is the beginning of step 2's coding process. This phase of work generates codes, not themes. Boyatzis (1998) defined a code as '[t]he most basic segment, or element, of the raw data or information that can be assessed in a meaningful way regarding the phenomenon' (p. 63). A code should be sufficiently well-defined and demarcated such that it does not overlap with other codes and should fit logically within a larger *coding framework* or *coding template* (sometimes also called *coding manuals*) that guides the coding process by outlining and defining the codes to be applied (Crabtree and Miller 1999; Attride-Stirling 2001; King 2004; Nowell et al. 2017). Similar to distinctions among themes, codes can be tied to more *semantic* or *latent* meanings (Braun and Clarke 2012), and the coding framework can be *inductive*, reflective of pertinent issues raised by the data alone, or *deductive*, guided by specific theories or theoretical frameworks (Attride-Stirling 2001; Braun and Clarke 2006). By recording how codes developed from observations and ideas, the researcher can begin the process of creating an audit trail to support the trustworthiness of a researcher's interpretations and analysis (Lincoln and Guba 1985; Nowell et al. 2017).

Once the coding framework or template is defined, researchers then apply the same codes to the entire data

set by labeling data extracts with relevant codes, making note of any potential patterns or connections between items that might inform subsequent theme development (Braun and Clarke 2006). Coding can be done manually or with the assistance of a computer program. Coded data extracts should include a large enough section of text to provide context for the extract. Also, a single extract can be labeled with multiple codes if relevant (Braun and Clarke 2006). Once the researcher has coded the entire data set, he or she can collate the data by code in preparation for step 3: the search for themes.

### Step 3: Searching for themes

The third step involves examination of the coded and collated data extracts to look for potential themes of broader significance (Braun and Clarke 2006). Braun and Clarke (2012) offer an analogy that, if your entire analysis is seen as a house, the individual codes are the bricks and tile, and themes are the walls and the roof. The process of theme identification – how those walls and roof are built – is fundamentally an active and interpretive process. Themes do not simply emerge from the data (Varpio et al. 2017); instead, themes are constructed by the researcher through analyzing, combining, comparing, and even graphically mapping how codes relate to one another. In inductive analysis, researchers derive themes expressly from the coded data, so the themes identified will be more closely linked to the original data and reflective of the entire data set (Braun and Clarke 2006). Conversely, in deductive analysis, predefined theories and/or theoretical frameworks will inform theme development, so these themes often focus more on a particular aspect of the data set or a specific question of interest (Braun and Clarke 2006).

When creating and organizing themes, thematic maps are useful for visually demonstrating cross-connections between concepts and among main themes and sub-themes (Braun and Clarke 2006). Figures 1 and 2 illustrate examples of thematic maps from our study. Themes should be independently meaningful but also 'work together to form a coherent whole—an analytic story' (Clarke and Braun 2014). Both Braun and Clarke (2006) and King (2004) recommended being inclusive at this stage of analysis. The researcher should make note of any and all themes of potential significance, whether or not they seem directly related to the study question, and regardless of the volume or quantity of data that falls under them. No defined

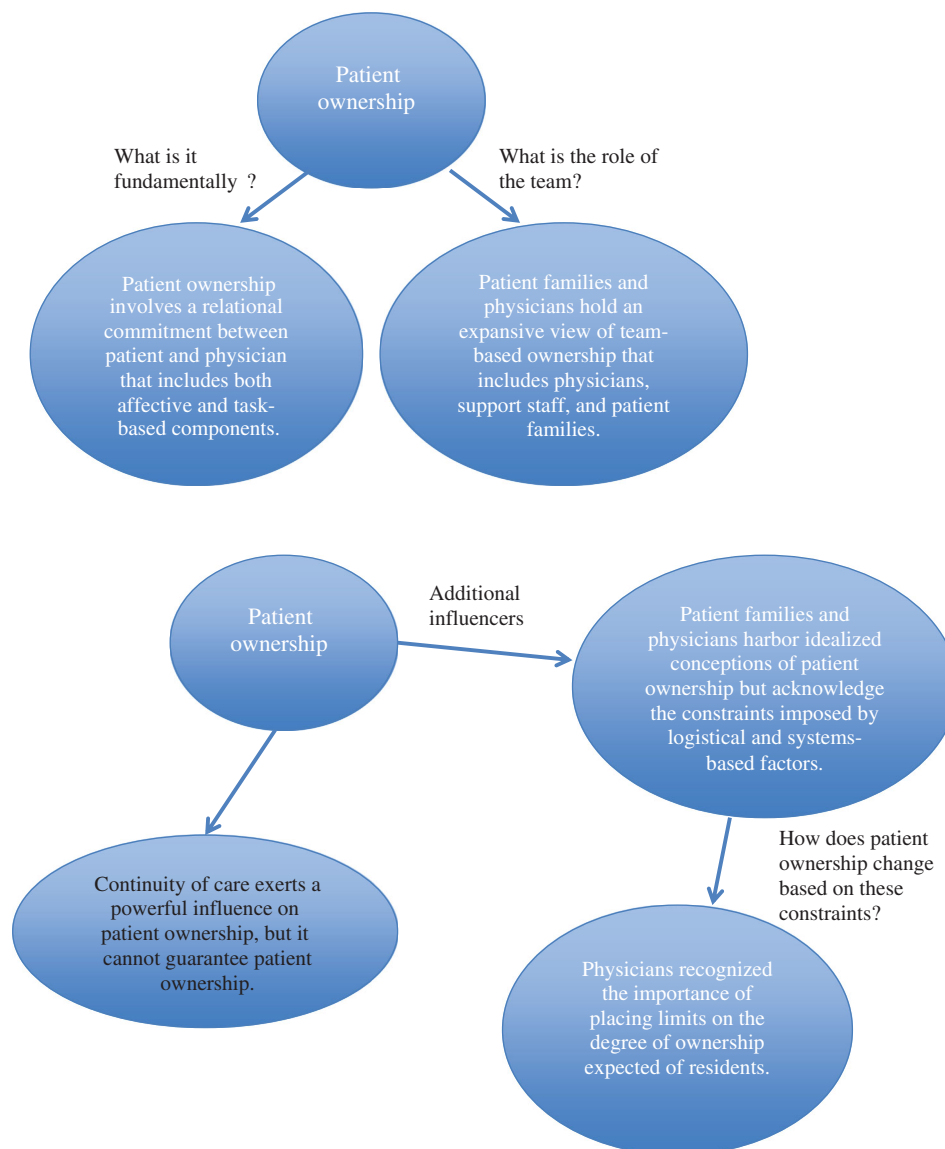


Figure 1. Macro-level thematic map for patient ownership project.

threshold exists for the amount of data that constitutes a theme. Researchers can even create a ‘miscellaneous’ theme to incorporate orphan codes that do not fit well within one’s existing thematic scheme (Braun and Clarke 2006). Important themes will provide significant links between data items and answer key aspects of research questions, but until reviewing the themes in step 4, the researcher cannot be certain of which themes will be kept, discarded, or otherwise modified in the final analysis.

#### Step 4: Reviewing themes

Braun and Clarke (2006) described step 4 as a two-level analytical process. In the first level of analysis, the researcher looks at coded data placed within each theme to ensure proper fit. He or she reviews all relevant codes and data extracts under each theme and asks: Does each theme have adequate supporting data? Are the data included coherent in supporting that theme? Are some themes too large or diverse? Data within each theme should have adequate commonality and coherence, and data between themes should be distinct enough to merit separation (Attride-Stirling 2001; Braun and Clarke 2006). At this point, data extracts can be re-sorted and themes modified to better reflect and capture coded data. Themes can

be added, combined, divided, or even discarded. This first level of analysis is complete when the researcher is confident that the revised thematic map adequately covers all of the coded data to be included in the final analysis (Braun and Clarke 2006). Throughout this process, researchers should keep detailed notes, or memos, regarding their thought processes and decisions made regarding how themes were developed, modified, and/or removed. Such memos can help researchers make connections between themes and also create an audit trail that bolsters the trustworthiness of their findings (Nowell et al. 2017).

Level two applies a similar set of questions to the themes in relation to the entire data set. The researcher here decides if individual themes fit meaningfully within the data set and whether the thematic map accurately and adequately represents the entire body of data (Braun and Clarke 2006). The thematic map should clearly demonstrate how themes interrelate and how they represent the question or construct of interest. To accomplish this task, the researcher re-reads the entire data set to reexamine themes and to re-code for additional data that falls under the themes that have been newly created or modified in this phase, then revises the thematic map accordingly (Braun and Clarke 2006).

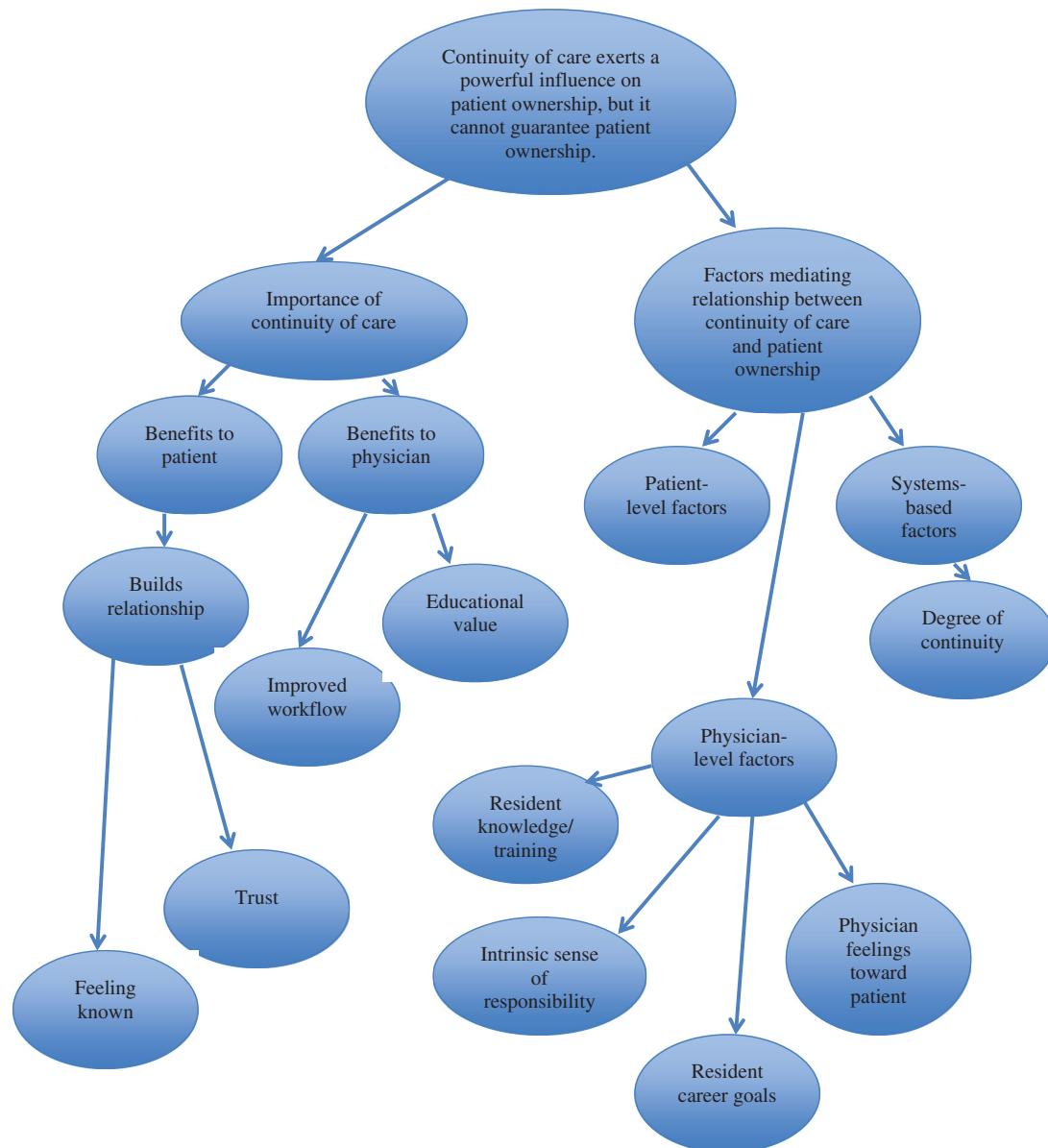


Figure 2. Micro-level thematic map focusing on highlighted theme.

Step 4 provides a clear example of the recursive nature of the entire process of thematic analysis. Re-reading and revising codes and themes are expected parts of the analytic process. While this iterative process could theoretically continue indefinitely, King (2004) and Braun and Clarke (2006) recommended that the revisions could cease once all data items that appear relevant to the study question have been incorporated into the coding scheme, themes are coherent, and additional refinements are not yielding substantial changes.

#### Step 5: Defining and naming themes

Once the thematic map has been refined, step 5 finds the researcher creating a definition and narrative description of each theme, including why it is important to the broader study question (Braun and Clarke 2006). The names of themes to be included in the final report are reviewed to ensure they are brief and adequately descriptive (Braun and Clarke 2006). The researcher then hones in on the most important aspect of each theme and which aspects of the data set it covers, creating a coherent narrative of how and why the coded data within each theme provide

unique insights, contribute to the overall understanding of larger questions, and interact with other themes (Braun and Clarke 2006). While addressing these questions, the researcher looks for areas of overlap between themes, identifies emergent sub-themes (which can be used both to provide more detailed accounts of themes and to describe hierarchies within the data), and clearly delimits the scope of what each theme entails or includes (Braun and Clarke 2006). This is a good time to select data extracts to be presented in the final report that illustrate key features of themes and to create narratives surrounding them that provide context to explain their importance to the broader story each theme tells (Braun and Clarke 2012).

#### Step 6: Producing the report/manuscript

This final step involves writing up the final analysis and description of findings (Braun and Clarke 2006). Elements of the writing process have already begun through the processes of note taking, describing of themes, and selection of representative data extracts in prior steps. In fact, King (2004) described the final step of presenting findings as a 'continuation' of the analysis and interpretation that



has already happened as opposed to a ‘separate stage’ (p. 267). The final report should move beyond mere description of codes and themes (King 2004). The report should weave a narrative that provides a clear, concise, and logical account not only *how* a researcher interprets the data, but also *why* his or her selection of themes and interpretation of the data are important and accurate (Braun and Clarke 2006, 2012). Using both narrative descriptions and representative data extracts (e.g. direct quotations from participants), the analysis should describe the data *and* provide an argument for why the researcher’s explanation richly and fully answers the research question (Braun and Clarke 2006). Any direct data extracts should contain adequate context to understand their meaning and be supported by interwoven textual description that explains their importance (Braun and Clarke 2012).

The discussion section can broaden the analysis by relating themes to larger questions, discussing implications of findings, and questioning the assumptions or preconditions that gave rise to the themes (Braun and Clarke 2006). Referencing related literature can also add to the strength of the analysis by building support for why particular themes were selected (Aronson 1995) and situating findings within the extant body of literature (Tuckett 2005). Given the flexibility that thematic analysis allows researchers regarding how to conduct data analysis, it is particularly important to identify choices (such as using inductive or deductive analysis in coding and identifying themes) and assumptions (including paradigmatic orientation) that underlie the analysis throughout the manuscript (Braun and Clarke 2006).

### **The advantages of thematic analysis**

As compared to many other qualitative methods, thematic analysis is relatively simple to learn and apply. Because it does not require the use of theory to inform analysis (i.e. it can be purely inductive) and because there are published descriptions and examples of the use of this analysis method, thematic analysis is quite accessible to less experienced researchers (King 2004; Braun and Clarke 2006; Nowell et al. 2017). At the same time, it is a powerful method for analyzing data that allows researchers to summarize, highlight key features of, and interpret a wide range of data sets. Furthermore, its methods are foundational to numerous other forms of qualitative analysis; in fact, Braun and Clarke (2006) argued that it should be the first method of qualitative analysis that researchers learn. Finally, and perhaps most importantly, thematic analysis offers researchers great flexibility with respect to: (a) the type of research questions it can address, from personal accounts of people’s experiences and understandings to broader constructs in various social contexts; (b) the type of data and documents examined; (c) the volume of data analyzed; (d) the choice of theoretical and/or epistemological framework applied; and (e) the ability to analyze data with an inductive, data-driven approach or a deductive, theory-driven approach (Clarke and Braun 2013).

### **The disadvantages of thematic analysis**

The flexibility that thematic analysis offers can also be seen as a drawback in that it contributes to a perception among some that it is not a rigorous method (Clarke and Braun 2013). If manuscripts don’t clearly state the paradigmatic orientation of the work and the role of theory in the analysis, thematic analysis risks being seen as a method that is applied broadly and never consistently. Additionally, given the flexible nature of this analysis method, it can be challenging for some researchers to determine which aspects of data to focus on and/or which theoretical or epistemological frameworks to use for their analysis (Braun and Clarke 2006). As with any analytic method, thematic analysis can be conducted poorly (see Pitfalls section below). However, a particular disadvantage with thematic analysis is that it has been more prone to inconsistent or improper use of terminology as compared to other methods with more well-defined and less flexible frameworks. These vagueries complicate appraisals of manuscripts that use, or claim to use, thematic analysis (Braun and Clarke 2006).

### **Pitfalls to avoid**

Braun and Clarke (2006) highlight several potential pitfalls to avoid when conducting thematic analysis, which broadly fall into three categories. First is a failure to adequately describe the assumptions that underlie the analysis. Thematic analysis will proceed in different ways if the method is used for inductive versus deductive analysis. It will also aim to achieve different goals if it is harnessed for a post-positivist investigation versus a constructivist investigation. These foundational aspects of the study should be explicitly stated, and then the researcher’s analysis must align itself to those foundations (Braun and Clarke 2006). Second is a lack of adequate analysis (Braun and Clarke 2006). Mere description of findings or a listing of data extracts or paraphrased responses does not constitute a thematic analysis; thematic analysis requires the researcher to engage in more interpretive work. For example, interview questions should not be used as themes, as this indicates a lack of analysis across the data set (Braun and Clarke 2006). Third is a weak analysis, in which the claims made in the report of findings are unconvincing or unfounded (Braun and Clarke 2006). This can occur when themes are overlapping, nonsensical, and/or lack internal consistency. Additionally, an analysis will be incomplete and unsupported if it lacks exemplary data extracts, fails to encompass all (or at least all of the most relevant parts) of the data set, elevates anecdotes or isolated events to the level of themes, or is contradicted by its data (Braun and Clarke 2006). Carefully following each of the six steps in the framework, which contain some built-in mechanisms for checking one’s work for internal consistency and completeness can help researchers using thematic analysis to steer clear of these pitfalls.

### **Conclusion**

Thematic analysis is clearly a powerful analytical method for qualitative research. We hope that this Guide not only makes clear what thematic analysis is and how it can be

used in different research traditions, but also lays bare the steps of the process (as described by Braun and Clarke) so that HPE researchers can use this method in their own work. Thematic analysis is a versatile qualitative data analysis method. We hope this description can help change our community's perception of thematic analysis. It is not a simple or unsophisticated data analysis; instead, it is a flexible and robust analysis method that can usefully help develop insights complex phenomena.

## Disclaimer

The views represented in this article are those of the authors alone and do not represent those of the United States Air Force, Department of Defense, or United States Government.

## Disclosure statement

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this article.

## Notes on contributors

**Dr. Michelle E. Kiger**, MD, is Assistant Professor, Department of Pediatrics, Uniformed Services University of the Health Sciences, Bethesda, Maryland, USA. She is a general pediatrician and the Military Element Program Director for the integrated Wright State University/Wright-Patterson Medical Center Pediatric Residency Program. Her research focuses on the concept of patient ownership and promotion of self-directed learning. She earned her MD from Vanderbilt University and is currently a PhD student in Health Professions Education at USU.

**Dr. Lara Varpio**, PhD, is Professor, Department of Medicine, Uniformed Services University of the Health Sciences, Bethesda, Maryland, USA. Her program of research investigates the many kinds of teams involved in medical education – e.g. inter professional care teams in clinical settings, medical education research and innovation unit teams, and scholarly collaboration teams. She uses qualitative methodologies and methods, along with theories from the social sciences and the humanities, to develop practical and theory-based evidence.

## ORCID

Michelle E. Kiger  <http://orcid.org/0000-0003-4506-6030>  
Lara Varpio  <http://orcid.org/0000-0002-1412-4341>

## References

- Aronson J. 1995. A pragmatic view of thematic analysis. *Qual Report*. 2(1):1–3.
- Attride-Stirling J. 2001. Thematic networks: an analytic tool for qualitative research. *Qual Res*. 1(3):385–405.
- Boyatzis RE. 1998. Transforming qualitative information: thematic analysis and code development. Thousand Oaks (CA): Sage.
- Braun V, Clarke V. 2006. Using thematic analysis in psychology. *Qual Res Psychol*. 3(2):77–101.
- Braun V, Clarke V. 2012. Thematic analysis. In: Cooper H, editor. *APA handbook of research methods in psychology*. Vol. 2, research designs. Washington (DC): American Psychological Association.
- Braun V, Clarke V. 2014. What can “thematic analysis” offer health and wellbeing researchers? *Int J Qual Stud Health Well-Being*. 9(1): 26152.
- Clarke V, Braun V. 2013. Teaching thematic analysis: overcoming challenges and developing strategies for effective learning. *Psychol*. 26(2):120–123.
- Clarke V, Braun V. 2014. Thematic analysis. In: Michalos AC, editor. *Encyclopaedia of quality of life and well-being research*. Dordrecht (Netherlands): Springer; p. 6626–6628.
- Clarke V, Braun V. 2017. Thematic analysis. *J Posit Psychol*. 12(3): 297–298.
- Crabtree B, Miller W. 1999. Using codes and code manuals: a template for organizing style of interpretation. In: Crabtree B, Miller W, editors. *Doing qualitative research* (2nd ed). Newbury Park (CA): Sage; p. 163–178.
- Glaser BG, Strauss AL. 1967. *The discovery of grounded theory: strategies for qualitative research*. Chicago (IL): Aldine.
- Joffe H. 2011. Thematic analysis. In: Harper D, Thompson AR, editors. *Qualitative methods in mental health and psychotherapy: a guide for students and practitioners*. Chichester (UK): John Wiley & Sons; p. 209–224.
- Kiger ME, Meyer HS, Hammond C, Miller KM, Dickey KJ, Hammond DV, Varpio L. 2019. Whose patient is this? A scoping review of patient ownership. *Acad Med*. 94:S95–S104.
- King N. 2004. Using templates in the thematic analysis of text. In: Cassell C, Symon G, editors. *Essential guide to qualitative methods in organizational research*. London (UK): Sage; p. 257–270.
- Lincoln Y, Guba EG. 1985. *Naturalistic inquiry*. Newbury Park (CA): Sage.
- Merton RK. 1975. Thematic analysis in science: notes on Holton's concept. *Sci Cult*. 188(4186):335–338.
- Nowell LS, Norris JM, White DE, Moules NJ. 2017. Thematic analysis: striving to meet the trustworthiness criteria. *Int J Qual Methods*. 16(1):160940691773384.
- Sandelowski M. 2010. What's in a name? Qualitative description revisited. *Res Nurs Health*. 33(1):77–84.
- Sandelowski M, Barroso J. 2003. Classifying the findings in qualitative studies. *Qual Health Res*. 13(7):905–923.
- Smith JA, Osborn M. 2003. Interpretative phenomenological analysis. In: Smith JA, editor. *Qualitative psychology: a practice guide to methods*. London (UK): Sage.
- Taylor J, Sims J, Haines TP. 2012. The influence of protection, palliation and costs on mobility optimization of residents in nursing homes: a thematic analysis of discourse. *Int J Nurs Stud*. 49(11):1364–1374.
- Thorne S. 2000. Data analysis in qualitative research. *Evid-Based Nurs*. 3(3):68–70.
- Tuckett AG. 2005. Applying thematic analysis theory to practice: a researcher's experience. *Contemp Nurse*. 19(1–2):75–87.
- Varpio L, Ajjawi R, Monrouxe LV, O'Brien BC, Rees CE. 2017. Shedding the cobra effect: problematising thematic emergence, triangulation, saturation and member checking. *Med Educ*. 51(1):40–50.
- Varpio L, Young M, Uijtdehaage S, Paradis E. 2019. Articulating the distinctions between *theory*, *theoretical framework*, and *conceptual framework*. *Acad Med*. Published Ahead of Print.
- Watling CJ, Lingard L. 2012. Grounded theory in medical education research: AMEE guide no. 70. *Med Teach*. 34(10):850–861.