Codes, Concepts and Categories, Oh My!
Building Your Skills in Qualitative Data Analysis

Alyssa Bogetz, MSW
Stanford University, abogetz@stanford.edu

Erika Abramson, MD, MS
Weill Cornell Medical College, err9009@med.cornell.edu

Hilary Haftel, MD, MHPE
University of Michigan, hils@med.umich.edu

Melissa Klein, MD
University of Connecticut, Melissa.Klein@cchm.org

Catherine Michelson, MD
Boston Medical Center, Catherine.Michelson@bmc.org

Arabella Simpkin, MD, MA
Harvard Medical School, arabella.simpkin@gmail.com

Su Ting Li, MD, MPH
University of California, Davis, sutli@ucdavis.edu
ACTIVITY 1: PRACTICE CODING A TRANSCRIPT

**Question:** What does it mean to “be well” during residency training?

**Respondent:** Chief Resident

Wellness is such a “hot topic” these days and for good reason. We all know residency is hard. You have little if any control over your schedule, you don’t get regular sleep or exercise, and at least in my case, you’re dealing with a lot of significant issues like death for the very first time. It’s kind of like residency is counter to wellness. But I think it’s because we see [wellness] as an end result. When I was a resident, I realized that I had to stop seeing wellness as a goal to be achieved, and instead, see it as a daily process of finding moments of peace, gratitude and joy. To me, this was/is wellness. Practicing wellness was deliberate - it had to be. I made the effort every day – whenever time allowed – to think of something I was grateful for. And even when things were crazy, I would find ways to incorporate wellness moments into my commute home, or sometimes, even at my front door right before entering, I would take a big deep breath, which really helped me find a sense of peace and grounding. Breathing really helps [laughs]! It does!

“Being well” is a process of finding moments to reconnect with yourself and I think also with others. I had a huge support system in my program – my co-residents were the best because they knew what I was going through. My wife, too, offered me stability and grounding in the midst of chaotic rotations. So I think that being well also means connecting with others. I think this relates to patients and families too and we can’t forget that. We have to find meaning in our work. I find moments of wellness in work when I get feedback from patients or families that they appreciate something I did or really felt like what we did had an impact on their child’s health. But I think you have to pay attention to this and listen for it each day. We often just go from one patient to the next and it’s exhausting, but wellness doesn’t just come to you. It’s not a goal you achieve. It’s a deliberate act that takes effort, attention, awareness and time. It’s a daily process of finding ways to connect with myself through gratitude, sometimes deep breathing– and to connect with patients, my friends and family. Wellness is finding meaning in my work and trusting that what I’m doing is making a difference. So it’s a mindset change. You don’t reach or find wellness, you practice it.
1. Read the transcript and develop a preliminary list of codes (10-15 total). Write your code list below.

2. In small groups of 3-4, discuss your codes. Agree on a single code list (i.e., codebook) and write this list below.

3. Using your agreed upon code list, individually code your transcript for a second time.
ACTIVITY 2: CREATE CATEGORIES AND THEMES

1. In your small groups of 3-4, review your coded transcripts and decide how to organize your codes into broader categories (10 or fewer). List these categories below, along with the individual codes representing them.

<table>
<thead>
<tr>
<th>Codes</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. Develop a list of 2-4 themes that link your categories together and provide a more “holistic” view of the data. Write your themes down in complete sentences. Share your themes with your table.

Bogetz A., Abramson E., Haftel H., Klein M., Michelson C., Simpkin A., Li, S. Codes, concepts and categories, oh my! Building your skills in qualitative data analysis. Association of Pediatric Program Directors (APPD), Anaheim, 2017.
## COMPARISON OF ANALYTIC APPROACHES

<table>
<thead>
<tr>
<th>Research Goal</th>
<th>Content Analysis</th>
<th>Thematic Analysis</th>
<th>Grounded Theory</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Develop categories • Document frequency of categories and/or codes to find meaning or compare across participants or groups</td>
<td>• Develop themes • Describe the “big picture” meaning of the data, often through interpretation of codes and categories</td>
<td>• Generate theory or a conceptual model about a social phenomenon or process</td>
<td></td>
</tr>
<tr>
<td><strong>Distinguishing Characteristics</strong></td>
<td>• Inductive or deductive • Allows for “cautious” quantification of data (i.e., codes, categories, words)</td>
<td>• Inductive or deductive • Allows for rich, detailed and nuanced account of data</td>
<td>• Inductive • Allows for deep exploration of a phenomenon to develop a related theory • Often involves systematic set of procedures (see below)</td>
</tr>
<tr>
<td><strong>Displaying Results</strong></td>
<td>• Table listing categories and representative quotations • Frequency counts of codes or categories • Quasi-statistical comparisons of codes between groups (e.g., faculty interview vs. resident interview)</td>
<td>• Table listing themes and representative quotations • In text quotations to illustrate themes and tell a “story”</td>
<td>• Figure that presents a conceptual framework or theory to highlight how ideas intersect</td>
</tr>
<tr>
<td><strong>How to Establish Credibility of Findings</strong></td>
<td>• Data triangulation - collect data from multiple sources • Peer debriefing with multiple analyzers - revise codes iteratively through discussion with team • Member check - ask participants to comment on accuracy of findings after analysis is complete • Reflexivity - acknowledge and minimize researcher bias in analytic process • Audit trail - keep a record of all analytic procedures and decision making</td>
<td></td>
<td>• Same as Content and Thematic plus: • Theoretical sampling - interviewing individuals who will add to the theory you are creating • Constant comparative analysis – reviewing and revising code list after each new interview</td>
</tr>
<tr>
<td><strong>Important Considerations</strong></td>
<td>• Good approach for survey data, more quantitative in nature • Less appropriate for open, explorative research</td>
<td>• Good approach for explorative research • Time consuming, labor intensive</td>
<td>• Requires knowledge of grounded theory approach • Resource intensive, often not practical for medical education research</td>
</tr>
</tbody>
</table>

Bogetz A., Abramson E., Haftel H., Klein M., Michelson C., Simpkin A., Li, S. Codes, concepts and categories, oh my! Building your skills in qualitative data analysis. Association of Pediatric Program Directors (APPD), Anaheim, 2017.
DATA MANAGEMENT AND ORGANIZATION

It is important to develop a data management strategy before you embark on a qualitative project. Many qualitative researchers enjoy manual coding and organization of data; however, qualitative software can be helpful for storing and organizing large databases and may be particularly beneficial if you are working with a team of investigators at multiple sites.

There are several qualitative analysis software packages available for purchase. These include:

1. Atlas.ti ($99/2 years)
2. Dedoose ($10.95/month)
3. Hyperresearch ($499 flat rate)
4. NVivo ($90)

Many of these programs offer similar features, so it is worth reviewing their websites and speaking with co-investigators or experienced qualitative researchers about their preferences before making your purchase.

Important features to consider when selecting your software are:

1. User-friendliness
2. Type of data the program will accept (e.g., can it handle text and image data?)
3. Reading and reviewing text (e.g., can it highlight quotations, can you search for specific text passages?)
4. Memo writing capability
5. Coding and categorization of data (e.g., can you easily apply codes to text or images, can you easily display, review and make changes to codes?)
6. Analysis and assessment of data (e.g., can you sort for specific codes, develop a concept map, make demographic comparisons with codes, assess interrater reliability?)
7. Team-based analysis allowance (i.e., can two or more researchers analyze the data together?)

If you do not wish to purchase software or are unable to, one effective and simple way to organize qualitative data is with a spreadsheet or word document. To organize your data, create a table that includes columns for the following: 1) codes, 2) code descriptions, 3) categories, and 3) representative quotations. A spreadsheet or word document is easy to refine and edit over time, and is an excellent way to stay organized. A separate file can be used for memos or additional reflections on the coding structure and emerging themes.
QUALITATIVE RESEARCH REFERENCES

Research Planning and Study Design


Data Analysis

Content Analysis:


Thematic Analysis:


Grounded Theory:

Exemplar Articles Using Each Analytical Approach

Content Analysis:


Thematic Analysis:


Grounded Theory:


Dissemination


Additional Materials

GLOSSARY OF TERMS

A **code** is a symbolic summary, usually a word or short phrase, that represents something that is larger than just the one piece of text (each code should be able to be applied to more than one piece of text).

A **category** is a word or phrase that groups codes together generally after coding is complete. It is helpful to think of a category as consisting of multiple codes.

A **theme** is a longer phrase or sentence that identifies or categorizes overarching principles seen in the data. Themes can be thought of as topic sentences describing and organizing how you present your data. Themes are more abstract ideas or patterns that often result from interpretation and abstraction of your data.

You **deductively** generate codes when you develop codes in advance of looking at the data, you know your specific research question a priori, and you do not change your codes. Codes generally come from existing literature or theory about your topic of interest.

You **inductively** generate codes when you enter into the analysis process with an open-ended research question and can allow yourself to develop codes from the data. You can change or modify the codes as you progress through the data, but have to modify your codebook as you make changes, and must return to your data several times to re-code with your final codebook.

You can generate codes through a **mixed inductive-deductive approach (abductive)** where you begin your analysis with a set number of pre-existing codes that you wish to use that come directly from the literature. You also iteratively change, add or modify your codebook over time, and incorporate new codes developed from the data itself.

**Content analysis** is a mechanism of counting or creating frequencies from qualitative data. Content analysis almost always uses a deductive coding approach. The researcher is able to count frequencies and show quantitative frequencies by groups. This is a good approach if you have a single question that was asked to all participants (i.e., open-ended survey questions, structured interview).

**Thematic analysis** means you analyze data for overarching ideas and not specific content. Thematic analysis usually involves iterative, ongoing review of the data. Instead of asking yourself as the researcher “what are participants talking about?” (which becomes a list or general topic) you want to ask “What are they expressing? What issues are they raising? What tensions are they experiencing?” that are likely to generate deeper meaning.

**Grounded theory** is a philosophical and highly systematic approach to data collection, coding and analysis that is iterative in its method (e.g., conduct one interview, do initial inductive coding, change interview guide, conduct a second interview, inductively add to codebook, change interview guide, conduct a third interview, etc). The goal is to develop an actual theory that is supported by the constant comparison of codes over the course of many iterations of the above process. Most people do not do grounded theory research despite what is written in the literature because it is time consuming and resource intensive. Typically, when you see that a researcher used “grounded theory,” what they mean is they iteratively and inductively developed a codebook for a thematic analysis, frequently employing a constant comparative approach.