Using Saturation to Estimate Qualitative Sample Sizes

Monique Hennink
Associate Professor
Rollins School of Public Health
Emory University

February 16, 2023

Bonnie Kaiser
Associate Professor
Anthropology/Global Health
University of California, San Diego
Origin of Saturation

• Developed in Grounded Theory (GT)
  o Theoretical saturation, embedded in an iterative process
  o Focuses on *data adequacy* not sample size per se
  o Applies during data collection

• Importance of Saturation
  o Reflects rigor & data validity
Broader Applications of Saturation

- **Saturation and Sample Size**
  - Data repetition, fewer issues arise
  - Focus on gauging sample size

- **Challenges**
  - Absence of inductive process
  - Unclear definition, how assessed & achieved
  - ‘Rubber stamping’ - vague references to ‘reaching saturation’
Knowledge Gaps

- No empirical guidelines on sample sizes for saturation
- No guidance on how to assess saturation or determine it was reached
- Variable definitions of saturation (e.g. saturation of what?)

- No guidance on estimating sample sizes for saturation *a priori*
  - Estimating saturation without data
  - For a research proposal
Empirical Tests of Saturation
Saturation in Interview Data
Figure 1: Timing of code development
### Table 2: Dimensions of codes by interview where code identified

<table>
<thead>
<tr>
<th>Code Name</th>
<th>Code Dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>By Interview 6</td>
</tr>
<tr>
<td>Feel Well</td>
<td>No illness (1)</td>
</tr>
<tr>
<td></td>
<td>Feel well (3)</td>
</tr>
<tr>
<td></td>
<td>Know viral load is stable (3)</td>
</tr>
<tr>
<td></td>
<td>Illness triggers clinic visit (3)</td>
</tr>
<tr>
<td></td>
<td>Have medication supply (4)</td>
</tr>
<tr>
<td>HIV Stigma</td>
<td>No stigma/conceal status (1)</td>
</tr>
<tr>
<td></td>
<td>Social stigma (1)</td>
</tr>
<tr>
<td></td>
<td>Witness others stigma (1)</td>
</tr>
<tr>
<td></td>
<td>Health treatment stigma (3)</td>
</tr>
<tr>
<td></td>
<td>Historic violence stigma (3)</td>
</tr>
<tr>
<td></td>
<td>Workplace stigma (3)</td>
</tr>
<tr>
<td></td>
<td>Friends avoid you (5)</td>
</tr>
<tr>
<td>Gay stigma</td>
<td>Gay stigma (9)</td>
</tr>
<tr>
<td></td>
<td>Education on stigma (9)</td>
</tr>
<tr>
<td>Stress of stigma</td>
<td>Stress of stigma (10)</td>
</tr>
<tr>
<td>Sexual disease stigma</td>
<td>Sexual disease stigma (12)</td>
</tr>
<tr>
<td>Health staff attitude</td>
<td>Health staff attitude (12)</td>
</tr>
<tr>
<td>Friends fear death</td>
<td>Friends fear death (12)</td>
</tr>
<tr>
<td>Self-stigma</td>
<td>Self-stigma (13)</td>
</tr>
<tr>
<td>Intimate partner stigma</td>
<td>Intimate partner stigma (17)</td>
</tr>
<tr>
<td>Family stigma</td>
<td>Family stigma (18)</td>
</tr>
<tr>
<td>Stigma of HIV death</td>
<td>Stigma of HIV death (22)</td>
</tr>
<tr>
<td>Job seekers disclosures</td>
<td>Job seekers disclosures (23)</td>
</tr>
<tr>
<td>Health insurance stigma</td>
<td>Health insurance stigma (23)</td>
</tr>
<tr>
<td>Perception of VA stigma</td>
<td>Perception of VA stigma (24)</td>
</tr>
</tbody>
</table>
Figure 4: Timing of code development versus timing of meaning saturation

- **Code**
  - Feel well
  - Enough Medication
  - Time
  - Disclosure
  - HIV stigma
  - Work commitments
  - Comfort with virus
  - Responsibility for health
  - Not a death sentence

- **Code saturation**
  - Code developed
  - Code meaning saturated

- **Interview number**
  - Range from 1 to 25
Saturation in Focus Group Data
Figure 1. Timing of code development and code saturation

*Y=Younger; O=Older; M=Male; F=Female

Note: Two inductive codes were developed in focus group 2. These codes were derived from questions in the discussion guide that were not probed in focus group 1 but were probed in focus group 2.

Legend:
- **Inductive**
- **Deductive**

<table>
<thead>
<tr>
<th>Focus group number</th>
<th>% of codes developed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>60</td>
</tr>
<tr>
<td>2</td>
<td>84</td>
</tr>
<tr>
<td>3</td>
<td>88</td>
</tr>
<tr>
<td>4</td>
<td>94</td>
</tr>
<tr>
<td>5</td>
<td>96</td>
</tr>
<tr>
<td>6</td>
<td>100</td>
</tr>
<tr>
<td>7</td>
<td>100</td>
</tr>
<tr>
<td>8</td>
<td>100</td>
</tr>
<tr>
<td>9</td>
<td>100</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
</tr>
</tbody>
</table>

FGD stratum:
- YM
- OF
- OM
- YF

*Y=Younger; O=Older; M=Male; F=Female
Meaning Saturation

Figure 4. Timing of first use of codes and their meaning saturation.

High prevalence codes
- Diabetes Cause
- US-Indian Food
- Exercise Barriers
- Exercise Perception
- Women’s Responsibility
- Cultural Expectations
- Mood

Low prevalence codes
- Food Taste
- Exercise Measures
- Exercise Facilities
- Ingredient Cost
- Exercise Gender
- Exercise Instructor
- Physical Appearance
- Work Success
- Exercise Pleasure
- Denial

*Non-italics = concrete code
*Italics = conceptual code

FGD NUMBER AND STRATUM
## Influence of Strata on Saturation

**Table 2. Examples of code dimensions identified across demographic strata of focus group discussions**

<table>
<thead>
<tr>
<th>Code</th>
<th>FGD 1 Strata: Young Men</th>
<th>FGD 2 Strata: Older Women</th>
<th>FGDs 3 &amp; 4 Strata: Older Men</th>
<th>FGD 5 Strata: Young Women</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Exercise Barriers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(concrete code)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Same issues repeated</td>
<td>Same issues repeated</td>
<td>Same issues repeated</td>
<td>Same issues repeated</td>
</tr>
<tr>
<td>Dimensions raised in specific strata</td>
<td>Little interest in physical appearance.</td>
<td>Little awareness of health benefits of exercise vs. weight loss.</td>
<td>Socializing valued over exercise.</td>
<td>Exercise routine is challenge.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>No family encouragement.</td>
<td>Need accompaniment.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Home exercise not effective.</td>
<td></td>
</tr>
<tr>
<td><strong>Mood</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(conceptual code)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dimensions raised across strata</td>
<td>Laziness to exercise</td>
<td>Same issue repeated</td>
<td>Same issue repeated</td>
<td>Same issue repeated</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dimensions raised in specific strata</td>
<td>Longing for family influences diet.</td>
<td>Apathy for diet once children grown</td>
<td>Stress eating influences diet.</td>
<td>Eating habits difficult to change.</td>
</tr>
<tr>
<td></td>
<td>Cravings for traditional foods.</td>
<td>Mental calm influences eating.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Satisfaction of food after long work hours.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Systematic Review of Saturation Tests
<table>
<thead>
<tr>
<th>Type of Approach</th>
<th>Description of Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Code Frequency Counts</strong></td>
<td>This approach involves reviewing each interview or focus group transcript and counting the number of new codes in each successive transcript or set of transcripts, until the frequency of new codes diminishes with few or no more codes identified. Several articles additionally randomized the order of data to assess the influence of sequential bias on saturation. Some articles added additional elements to the code frequency counts, such as batch comparison, a stopping criterion or saturation of higher order groupings of data, as outlined below.</td>
</tr>
<tr>
<td><strong>Comparative Method</strong></td>
<td>This approach adds a more structured comparison to the code frequency count approach above. It involves reviewing data in pre-determined batches, such as quartiles of data (instead of reviewing each interview separately) and listing all new codes in a saturation table for each batch of data. The subsequent quartile of data is then reviewed and compared to the first quartile to determine any new codes, this comparison of data batches continues until few or no new codes are identified, whereby saturation is achieved.</td>
</tr>
<tr>
<td><strong>Stopping Criterion</strong></td>
<td>This approach adds a stopping criterion to the code frequency count approach above. It involves reviewing an initial sample of interviews (e.g. 6 interviews) or focus groups to identify new codes, and using a pre-determined stopping criterion, which is usually the number of consecutive interviews/groups after the initial sample where no new codes are identified in the sample (e.g. 2 or 3 interviews with no new codes). Saturation is reached when no new codes are identified after the stopping criterion of x interviews after the initial sample, or the number of new codes is under a predetermined threshold (e.g. &lt;5%). In other studies, the stopping criterion was based on repetitions of a code, such as 3 or 5 instances of a particular code or theme were identified.</td>
</tr>
<tr>
<td><strong>High-Order Groupings</strong></td>
<td>This approach uses a higher order grouping of codes in the code frequency count approach above. It involves counting higher-order groupings of codes such as meta-themes, salient themes or categories. For example, Coenen et al (2012) counted conceptual categories. Hagaman et al (2016) counted codes to determine the most prevalent codes in the data set, then randomized the interview order via bootstrapping to determine the average number of interviews needed to identify the most prevalent codes in data. Weller et al (2018) focused on identifying saturation for the most salient items in data.</td>
</tr>
<tr>
<td><strong>Code Meaning</strong></td>
<td>This approach does not focus on counting codes as the basis for determining saturation (as used in the approaches above), instead achieving a full understanding of codes is the indicator of saturation. It involves reviewing an interview and noting each issue (or code) identified, then in subsequent interviews identifying whether any new aspects, dimensions, or nuances of that code are identified, until nothing new is identified and the code has reached saturation. Codes may reach saturation at a different point in the data set.</td>
</tr>
</tbody>
</table>

Table 3: Strategies to Assess Saturation in Empirical Tests
<table>
<thead>
<tr>
<th>Data Application</th>
<th>Strategy to Assess Saturation</th>
<th>Parameters and Assumptions</th>
<th>Suggested formula for saturation</th>
</tr>
</thead>
</table>
| **Fofana et al. (2020) PLOS ONE** | Uses set theory and partial least squares regression to estimate saturation | \( X_j \) is the vector of the number of times each theme is coded in the \( j \)-th interview \( B_{PLS} \) is the vector of regression coefficients \( E \) is the matrix of residuals | \( (X_1, \ldots, X_n) = (X_1, \ldots, X_n) \)
| **Fugard & Potts (2015) Int. J. Soc. Res. Methodology** | Uses negative binomial probability distribution to estimate sample needed to reach a certain power (e.g., 80% probability to identify a theme) based on several parameters | Assumes random sample. Estimates sample size based on population theme prevalence (known probability of issue/theme in the population of interest) of least prevalent theme, desired number of instances in the data, and desired power. | Various outcomes provided based on values for model inputs |
| **Galvin (2014) J. Building Engineering** | Uses binomial distribution to answer 5 research questions; the most relevant is RQ3: How many interviews to have 95% probability of theme emerging? | Assumes random sample | \( n = \frac{\ln(1-P)}{\ln(1-\alpha)} \)
| **Lowe et al. (2018) Field Methods** | Develops saturation index using generalized estimating equations | \( R = \text{prevalence of a theme in population} \) \( P = \text{particular saturation} \) \( n = \# \text{observations} \) Accounts for statistical dependency between observations and likelihood of researcher identifying theme. Assumes order of observations does not influence themes identified. Assumes random sample | \( n = \frac{P(1-P)}{R(P-1)} \)
| **Rowlands et al. (2015) J. Computer Inf. Systems** | Calculate thematic saturation using lognormal distribution with chosen confidence level | Based on concept analysis using Leximancer program. \( X^{*} \) is the geometric mean from the lognormal fit \( s^{*} \) is the multiplicative standard deviation from the lognormal fit | For 95% confidence lognormal expression 
\( X^{*} \pm s^{*} \times z_{0.975} \)
| **Van Rijnsoever et al. (2017) PLOS ONE** | Uses simulations based on lognormal distribution and 11 parameters | Accounts for random and purposive samples, as well as minimal and maximal information from observations. | Various outcomes provided based on values for model inputs |
Saturation in Interview Data

Sample Size for Saturation in Empirical Tests with Interview Data

- Meaning saturation
- Meta-themes
- Themes
- Code saturation
- Inductive approach
- Deductive approach
- Randomized order
- Actual order of interviews

Study dataset (listed by author)
Saturation in Focus Group Data

Sample Size for Saturation in Empirical Tests with Focus Group Discussion Data

Number of FGDs for saturation

Study dataset (listed by author):
- Hancock
- Hennink
- Guest
- Coenen
- Young
- Morse

Hancock: 4
Hennink: 8
Guest: 2
Coenen: 6
Young: 1
Morse: 18

Meaning saturation
Code saturation
Implications of Findings

• Provide empirical guidance on sample sizes for saturation as start point

• Give evidence to refute critiques of “small samples”

• Focus sample size estimation on data not n’s

• Encourage more informed critiques of qualitative sample sizes & justifications

• Provides researchers with strategies to assess saturation to encourage transparency
Estimating Sample Sizes
Parameters Influencing Saturation

- Study Purpose
- Study Population
- Sampling Strategy
- Data Quality
- Type of Codes
- Saturation Goal
- Degree of Saturation
- Stratification

Estimate Sample Size for Saturation

Assess Saturation in Data Collection

Source: Hennink (2017, 2020)
Figure 6: Parameters of saturation and sample sizes.
A clinic director recently implemented a new electronic health record (EHR) alert that aims to increase delivery of an evidence-based practice. The director is interested in understanding doctors’ experiences of the EHR alert training. They aim to recruit doctors from their medium-sized clinic, which serves a suburban area. The goal of qualitative analysis is to identify potential issues that could be used to improve the EHR alert and/or training. The clinic director has not conducted qualitative research previously.
Case Study

A group of researchers is interested in studying why therapists discontinue use of evidence-based practices (EBPs). They aim to recruit therapists who have used various EBPs in children’s mental healthcare agencies across a large county. Therapists working in the agencies are diverse in terms of age, race/ethnicity, EBPs employed, work site and hours, and (somewhat) education. As part of an explanatory (QUAN --> qual) mixed-method design, qualitative analysis sought to explain quant findings by describing and comparing therapist experiences and perceptions.

Adapted from Lau et al. (2020)
Further Research

- Parameters influencing saturation
- Inductive data collection
- Less homogenous study populations
- Different types of data, code styles, saturation
Reference Articles

