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)*
    - Theoretical saturation, embedded in an iterative process
    - Focuses on *data adequacy* not sample size per se
    - Applies during data collection

- Importance of Saturation
    - 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 <u>assess</u> saturation or determine it was <u>reached</u>
- Variable definitions of saturation (e.g. saturation of what?)

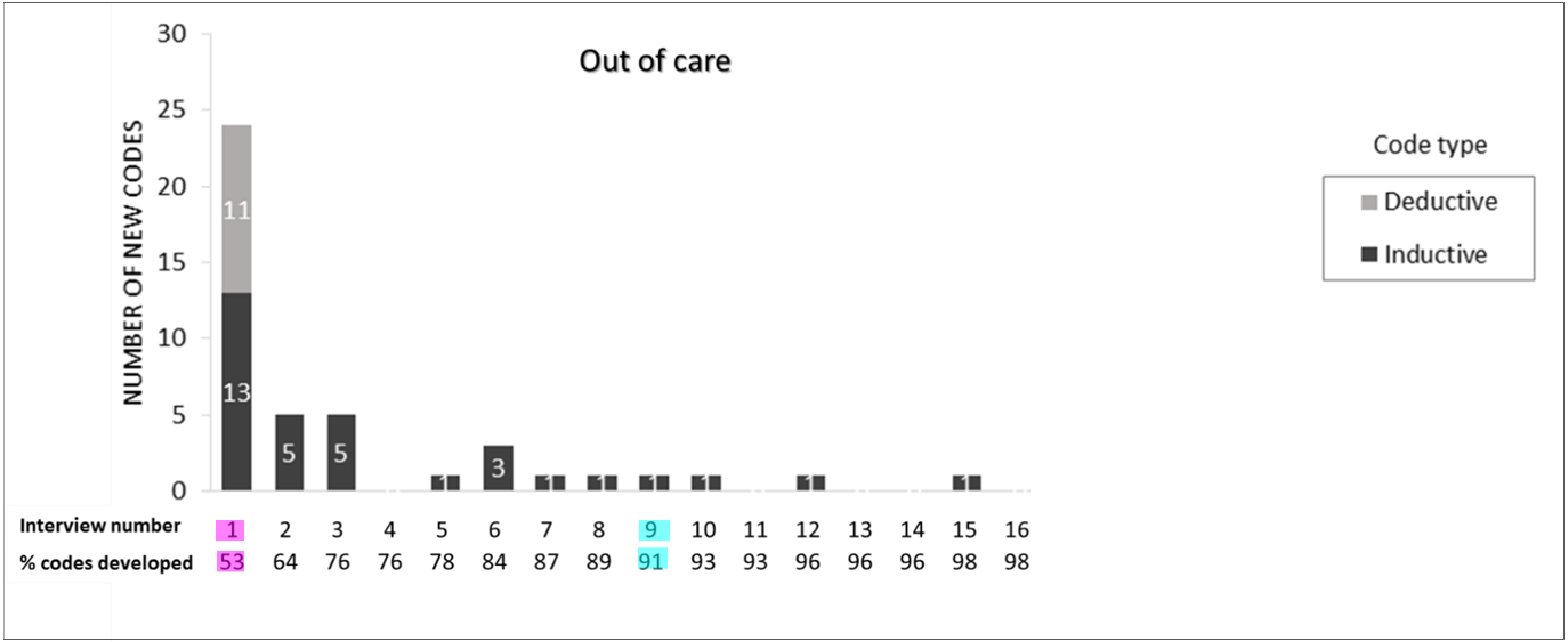
- No guidance on estimating sample sizes for saturation <u>a priori</u>
  - Estimating saturation without data
  - For a research proposal

Empirical Tests of Saturation

Saturation in Interview Data

Code Saturation

Figure 1: Timing of code development



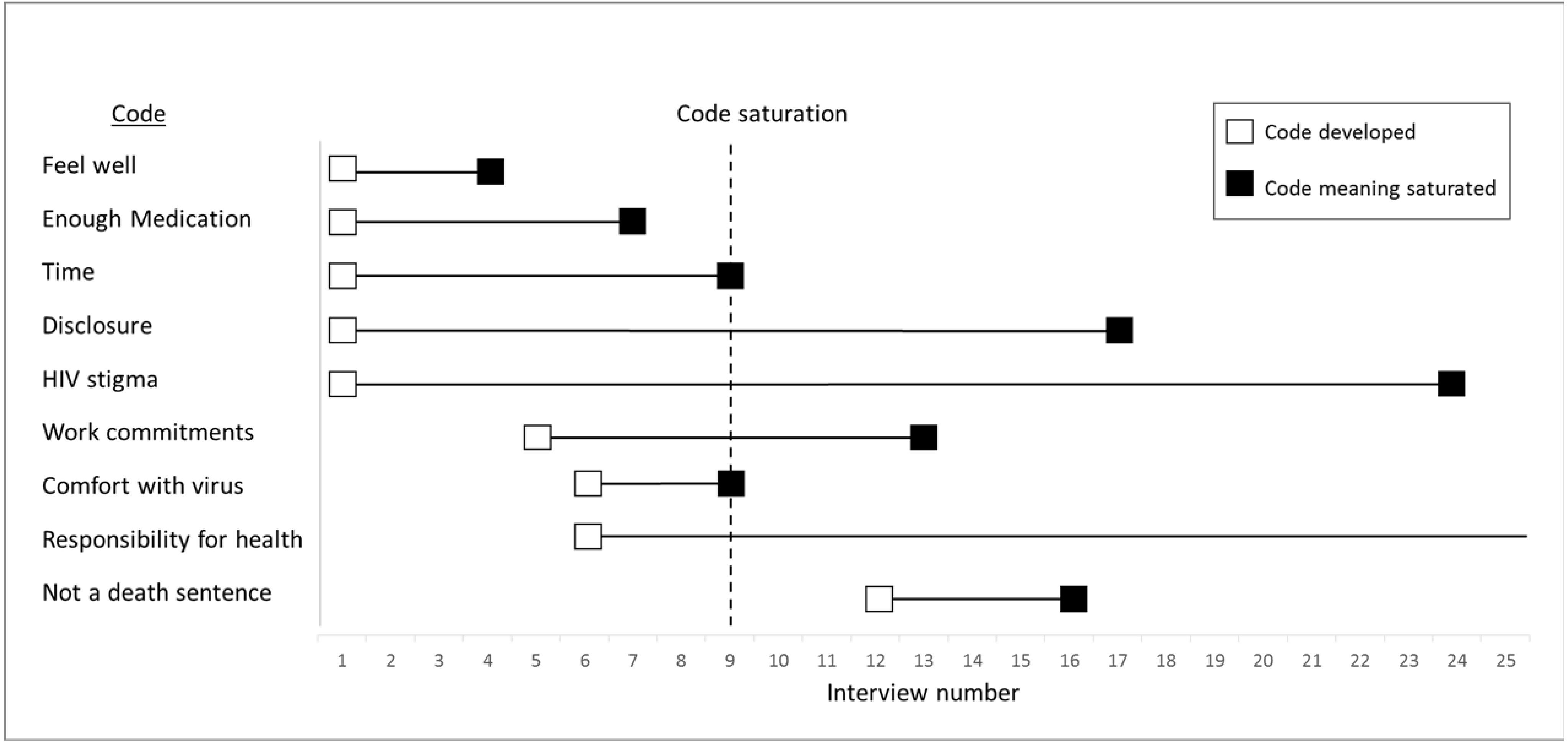
Meaning Saturation

Table 2: Dimensions of codes by interview where code identified

Code Name	Code Dimensions (# denotes interview number where identified)			
	By Interview 6	**By Interview 9**	**By Interview 12**	**After Interview 12**
Feel Well	No illness (1) Feel well (3) Know viral load is stable (3) Illness triggers clinic visit (3) Have medication supply (4)	None	None	None

Meaning Saturation

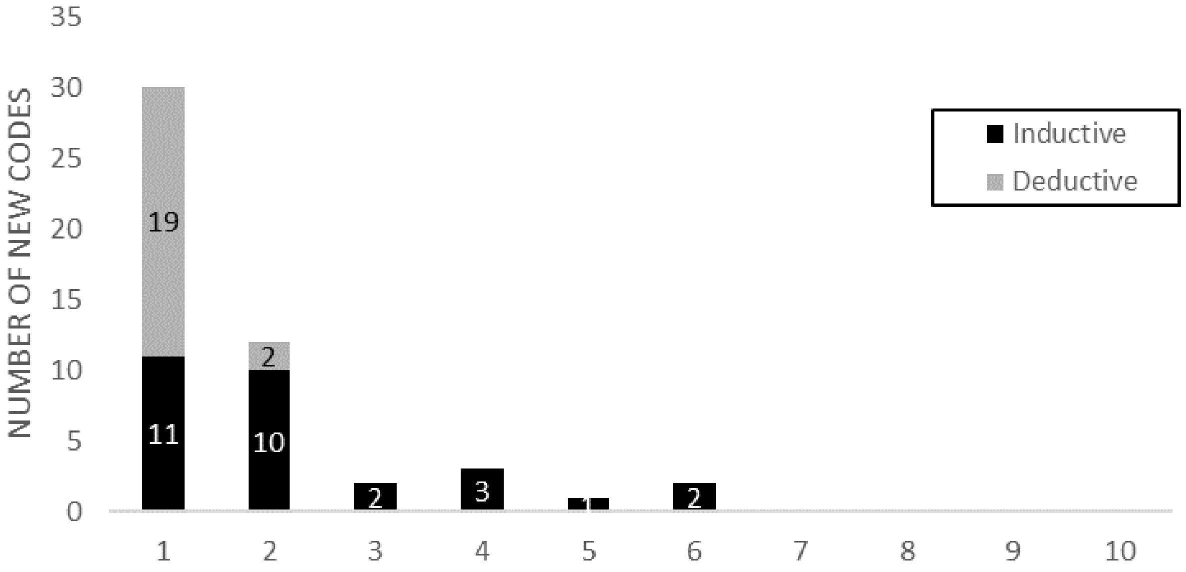
Figure 4: Timing of code development versus timing of meaning saturation



Saturation in Focus Group Data

Code Saturation

Figure 1. Timing of code development and code saturation

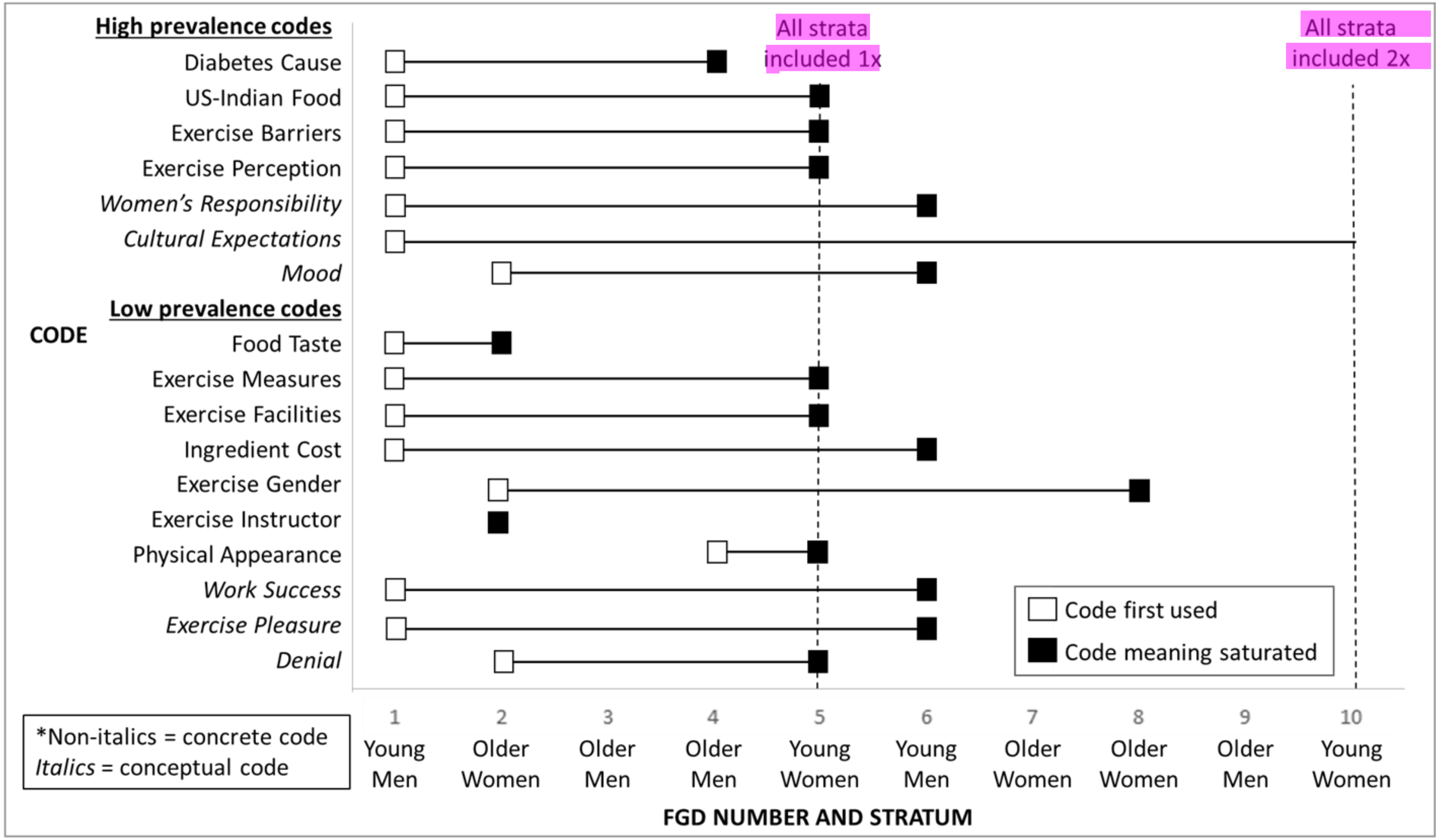


Focus group number	1	2	3	4	5	6	7	8	9	10
% of codes developed	60	84	88	94	96	100	100	100	100	100
FGD stratum*	YM	OF	OM	OM	YF	YM	OF	OF	OM	YF

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

Meaning Saturation

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



Influence of Strata on Saturation

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

Code	**FGD 1** **Strata: Young Men**	**FGD 2** **Strata: Older Women**	**FGDs 3 & 4** **Strata: Older Men**	**FGD 5** **Strata: Young Women**
Exercise Barriers (concrete code)				
Dimensions raised across strata	Lack of <u>time</u> for exercise. Lack of exercise <u>facilities.</u> Exercise <u>conflicts</u> with family time. <u>Cost</u> of exercise activities.	Same issues repeated	Same issues repeated	Same issues repeated
Dimensions raised in specific strata	Little interest in <u>physical appearance</u>. <u>Education prioritized</u> over physical activity.	Little <u>awareness</u> of health benefits of exercise vs. weight loss.	<u>Socializing</u> valued over exercise. Exercise <u>routine</u> is challenge. <u>Weather</u> limits outdoor exercise.	No family <u>encouragement</u>. Need <u>accompaniment</u>. <u>Home exercise</u> not effective.
Mood (conceptual code)				
Dimensions raised Across strata.	<u>Laziness</u> to exercise	Same issue repeated	Same issue repeated	Same issue repeated
Dimensions raised in specific strata	<u>Longing</u> for family influences diet. <u>Cravings</u> for traditional foods. <u>Satisfaction</u> of food after long work hours.	<u>Apathy</u> for diet once children grown	<u>Stress</u> eating influences diet. <u>Mental calm</u> influences eating.	Eating <u>habits</u> difficult to change.

Systematic Review of Saturation Tests

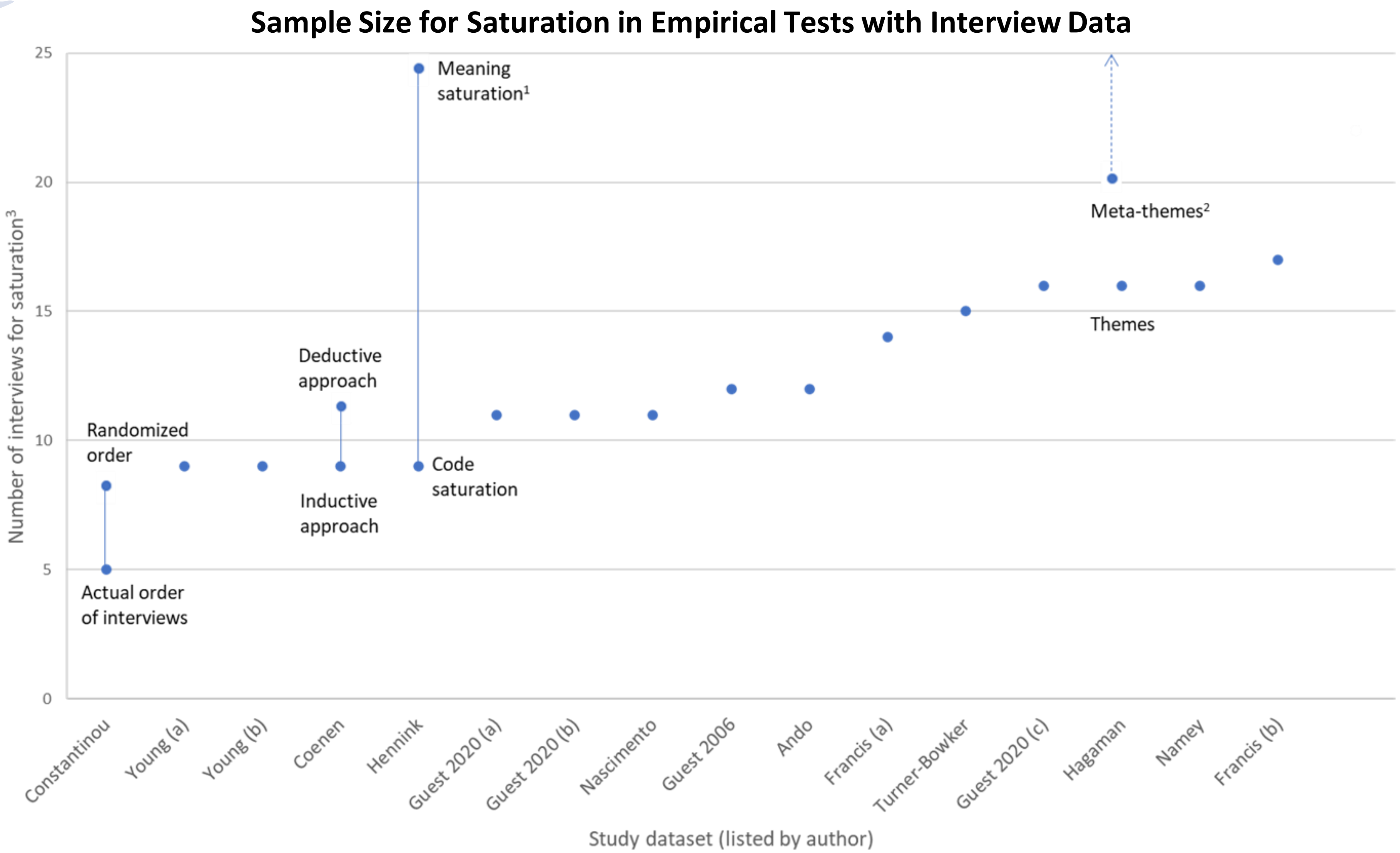
Table 3: Strategies to Assess Saturation in Empirical Tests

Type of Approach	Description of Approach
Code Frequency Counts	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.
Comparative Method	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.
Stopping Criterion	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. <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.
High-Order Groupings	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.
Code Meaning	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.

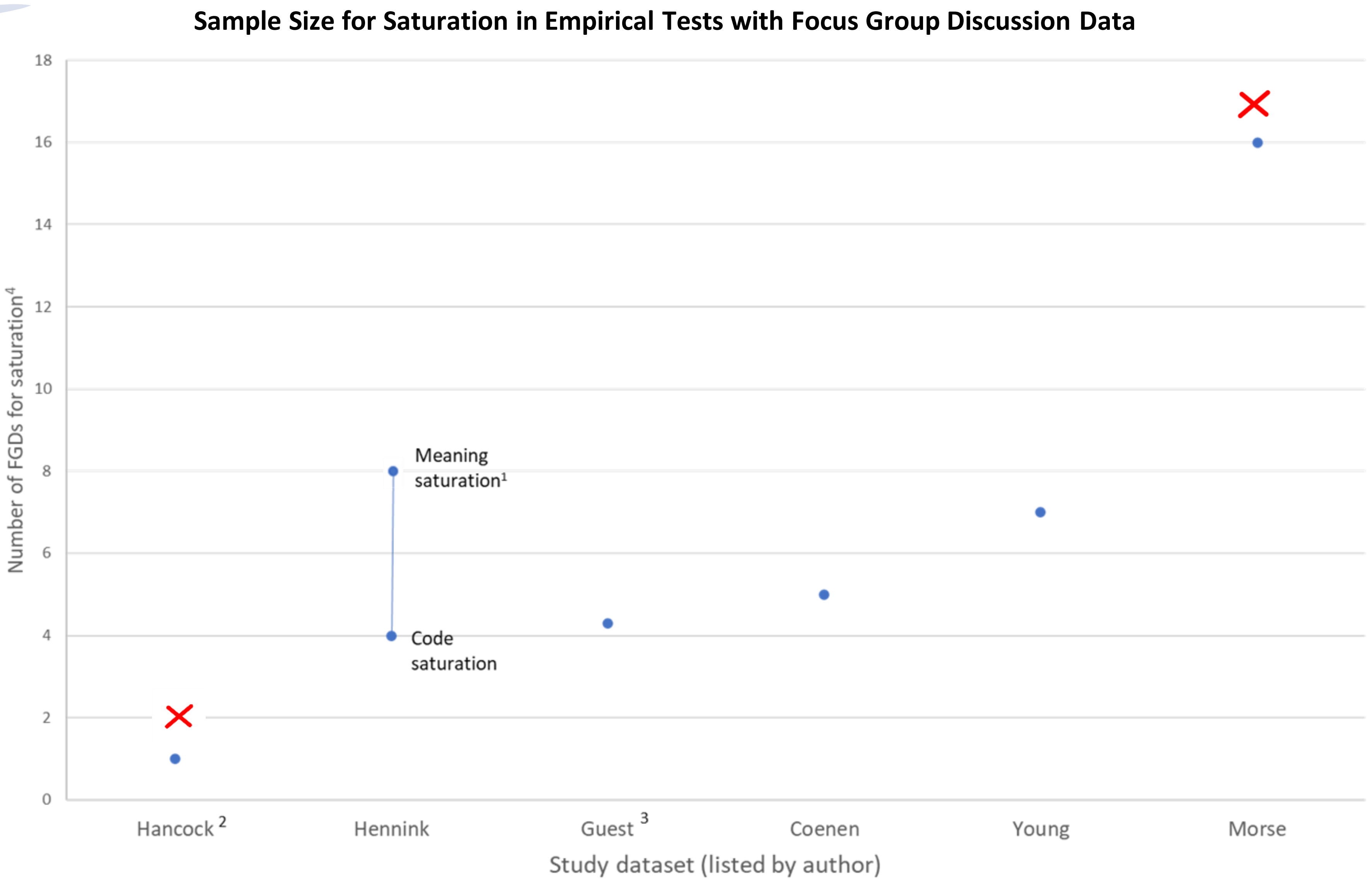
Table 2: Strategies to Assess Saturation through Statistical Tests

	Data Application	Strategy to Assess Saturation	Parameters and Assumptions	Suggested formul saturation
Fofana et al. (2020) *PLOS ONE*	Statistical model tested on empirical dataset of interviews (n=12)	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	$(X_{j+1} \dots X_n) = (X_1 \dots$
Fugard & Potts (2015) *Int. J. Soc. Res. Methodology*	Hypothetical model based on interviews but not tested on empirical data	Uses negative binomial probability distribution to estimate sample needed to reach a certain power (eg, 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 o of values for mod
Galvin (2014) *J. Building Engineering*	Hypothetical model based on interviews but not tested on empirical data	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 P = probability theme arising in interview R = proportion of theme in population n = # interviews	$n = \frac{ln(1-P)}{ln(1-R)}$
Lowe et al. (2018) *Field Methods*	Statistical model tested on empirical datasets including literature surveys (n=25), focus groups (n=3), and interviews (n=11)	Develops saturation index using generalized estimating equations	R = prevalence of a theme in population P = particular saturation n = # 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	$n = \frac{P(R-1)}{R(P-1)}$
Rowlands et al. (2015) *J. Computer Inf. Systems*	Statistical model tested on empirical data of interviews (3 studies: n = 30, 30, 24)	Calculate thematic saturation using lognormal distribution with chosen confidence level	Based on concept analysis using Leximancer program. $\overline{X}^*$ is the geometric mean from the lognormal fit s^* is the multiplicative standard deviation from the lognormal fit	For 95% confiden lognormal expres $=\overline{X}^* * (s^*)^2$
Van Rijnsoever et al. (2017) *PLOS ONE*	Hypothetical model based on various data types (e.g., interviews, focus groups, documents) but not tested on empirical data	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 o of values for mod

Saturation in Interview Data



Saturation in Focus Group Data



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

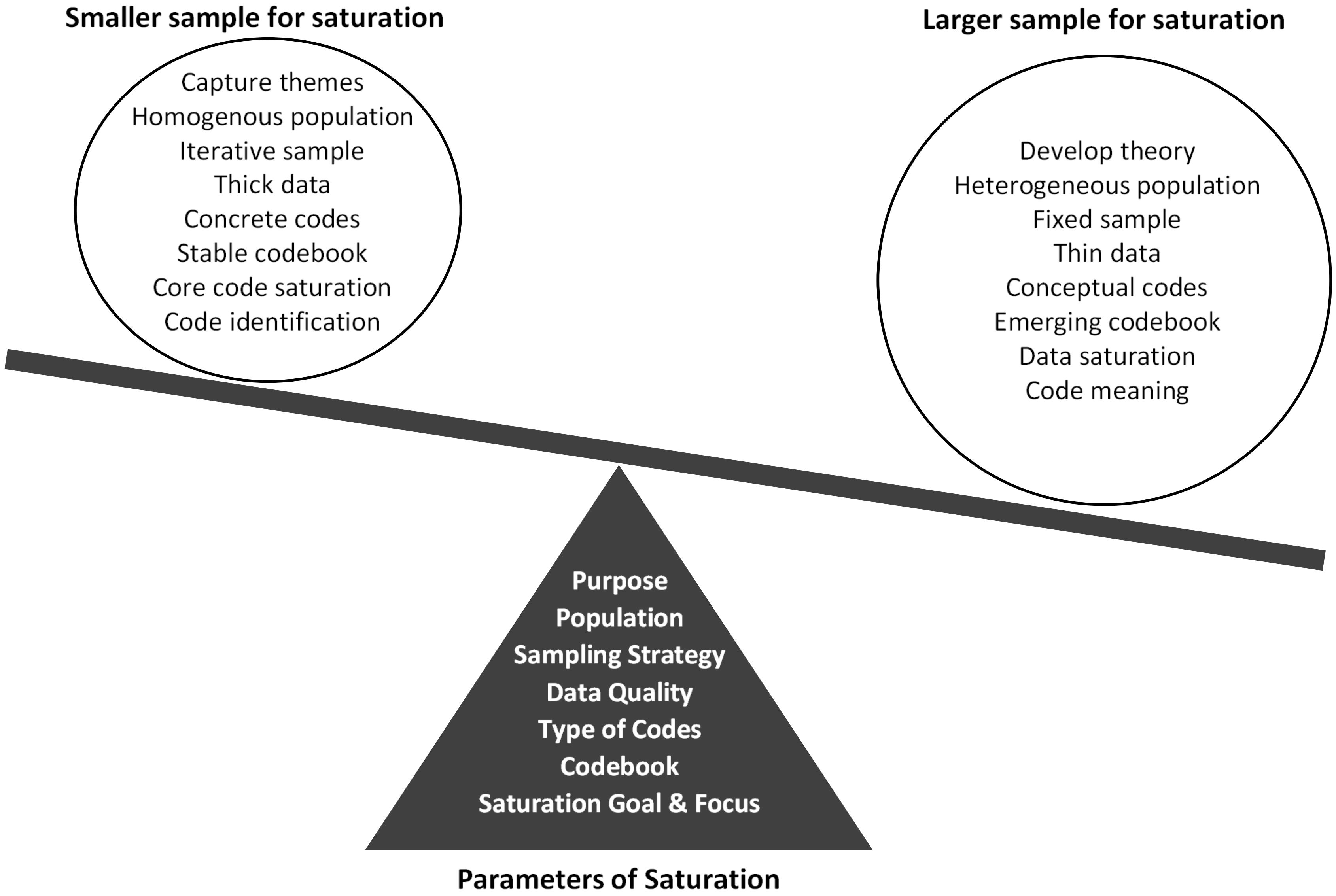
Parameters Influencing Saturation

- 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.



Case Study

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

- Hennink, M., and Kaiser, B. (2022). Sample Sizes for Saturation in Qualitative Research: A Systematic Review of Empirical Tests. *Social Science and Medicine*. 292, (2022) 114523

- Hennink, M., Kaiser, B., and Weber, MB. (2019) What Influences Saturation? Estimating Sample Sizes in Focus Group Research" *Qualitative Health Research*. Volume 29, Issue 10.

- Hennink, M., Kaiser, B. and Marconi, V. (2016) Code Saturation vs. Meaning Saturation: How many interviews are enough? *Qualitative Health Research. Volume 27(4), 591-608.*