RESEARCH · QUALITATIVE RESEARCH

Saturation in Qualitative Research: How Many Interviews Are Enough?

What saturation means in qualitative research and how to decide how many interviews or focus groups are enough, covering the origins of the concept, its different meanings, empirical studies of when saturation occurs, information power, stopping rules, the critique of saturation, and practical guidance for planning and reporting sample size.

ModQu research article · 2,894 words · 25 references (APA 7th edition)

"How many interviews do we need?" is one of the most common questions in qualitative research, and one of the hardest to answer. In quantitative research, sample size follows from a calculation of statistical power. Qualitative research has no equivalent formula. The most common answer is that data collection should continue until **saturation**: the point at which further data add little or nothing new. Saturation is cited in a large share of qualitative studies as the justification for their sample size. Yet the concept is used in several different senses, it is often claimed without evidence, and some researchers argue that it does not fit many forms of qualitative analysis at all. This article explains where the concept came from, what it means, what the evidence says about when it is reached, and how to plan and justify sample size in practice.

Origins: theoretical saturation

The idea comes from grounded theory. Glaser and Strauss (1967) described **theoretical saturation** as the point at which no additional data are being found whereby the researcher can develop properties of a category. As the researcher sees similar instances over and over again, they become confident that the category is saturated. Saturation in this original sense applies to categories in an emerging theory, not to the dataset as a whole, and it is linked to **theoretical sampling**: the researcher deliberately seeks new cases that might reveal new properties, rather than simply collecting more of the same.

Charmaz (2014) emphasised that saturation in this sense means that the properties of the theoretical categories are fully developed, not merely that the same stories keep being repeated. Dey (1999) cautioned that saturation is a judgement made by the researcher, and suggested that "theoretical sufficiency" might be a more accurate term, since it is never possible to know that nothing new would emerge with more data.

Many meanings

As the concept spread beyond grounded theory, its meaning broadened. Saunders et al. (2018) identified four main models in use:

- **Theoretical saturation**, the original grounded theory sense, concerning the development of categories and theory.
- **Inductive thematic saturation**, concerning the emergence of new codes or themes during analysis.
- **A priori thematic saturation**, concerning the extent to which pre-defined codes or theoretical concepts are represented in the data.
- **Data saturation**, concerning the point at which new data repeat what has already been expressed.

These are different things. A study can reach data saturation, in which participants keep telling similar stories, without reaching theoretical saturation, in which the researcher understands the variations and conditions of each category. Saunders et al. (2018) recommended that researchers state which sense they mean and how it fits their methodology. O'Reilly and Parker (2013) argued that saturation had been adopted too uncritically as a universal marker of quality, and that the term should only be used where it fits the approach.

What the evidence shows

Several studies have examined empirically when saturation occurs.

Guest et al. (2006) analysed 60 interviews with women in two West African countries, coding them in batches of six and tracking when new codes appeared. They found that most codes, including all the most frequently used, had appeared within the first 12 interviews, and that the basic elements of the main themes were present after six. They cautioned that this finding applied to a fairly homogeneous group and a focused set of questions.

Hennink et al. (2017) distinguished **code saturation**, the point at which no new issues are identified, from **meaning saturation**, the point at which the issues are fully understood. In their study of 25 interviews, code saturation was reached after about nine interviews, but meaning saturation required between 16 and 24, depending on the code. Some codes, particularly complex or less common ones, needed more data to understand than to identify.

For focus groups, Guest, Namey and McKenna (2017) found that two to three groups were enough to capture about 80% of themes, including the most prevalent, and that three to six groups captured about 90%.

Hennink and Kaiser (2022) reviewed 23 empirical studies that tested saturation and found that saturation was typically reached with between 9 and 17 interviews, or between four and eight focus groups, in studies with relatively homogeneous populations and narrowly defined objectives. Broader studies and more diverse populations would need more.

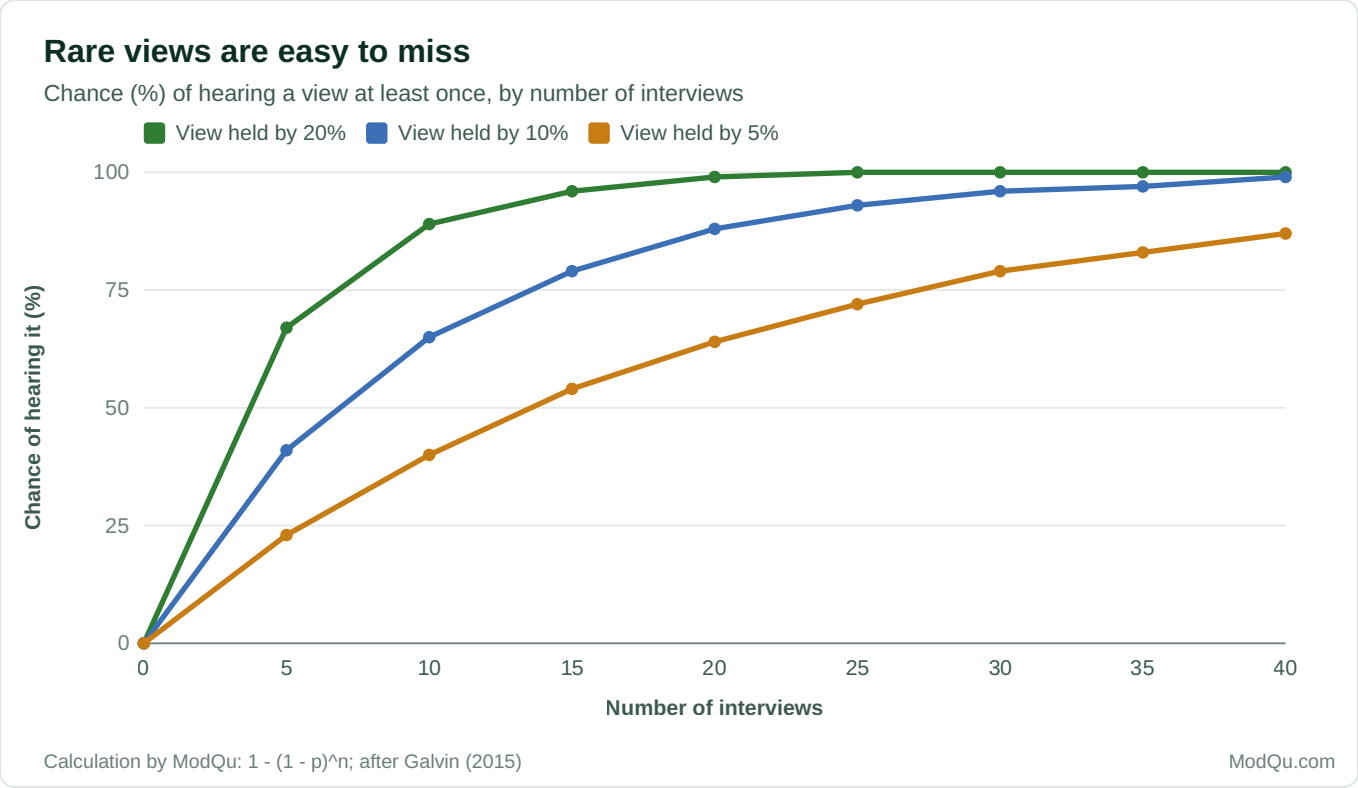


Figure 1. The chance of hearing a view at least once, by number of interviews and how common the view is. Illustrative calculation by ModQu assuming independent random selection, following the logic of Galvin (2015). Purposive samples differ, but rare views are still easy to miss.

Simulation and probability studies add a further perspective. van Rijnsoever (2017) simulated sampling from populations with different numbers of codes and found that the sample size needed depends on how many codes exist and how likely each is to be observed, and that purposive sampling reaches saturation faster than random sampling. Galvin (2015) used simple probability reasoning to show how many interviews are needed to be confident of encountering a view held by a given share of the population. The chart above illustrates this logic: a view held by 20% of a population is very likely to appear at least once within 15 interviews, but a view held by 5% may well be missed even after 30.

The logic has limits, because qualitative samples are not random and the aim is not to estimate prevalence. It is nevertheless a useful reminder that rare experiences are easily missed, and that if they matter to the research question, sampling should seek them out deliberately.

What researchers actually do

Studies of published research suggest that saturation is more often claimed than demonstrated. Mason (2010) examined 560 doctoral theses based on qualitative interviews and found that sample sizes clustered at round numbers, such as 20 and 30, which suggests that they were set in advance rather than determined by saturation. Vasileiou et al. (2018) reviewed 214 interview-based articles in health journals over 15 years and found that most did not justify their sample size at all, and that those that did usually cited saturation without explaining how it had been judged. Morse (1995) had made a similar observation much earlier, noting that saturation was frequently claimed but rarely described.

Stopping rules

To make saturation more transparent, several researchers have proposed explicit **stopping rules**. Francis et al. (2010) proposed that researchers specify an initial analysis sample, such as 10 interviews, and a stopping criterion, such as three consecutive further interviews with no new themes. Data collection stops when the criterion is met. Guest et al. (2020) proposed a related method in which a **base size** of interviews is analysed to establish the initial set of themes, and then subsequent **runs** of interviews are examined. Saturation is reached when a run adds new information below a chosen threshold, such as 5% of the themes in the base.

Lowe et al. (2018) argued that some commonly used stopping rules can end data collection too early, and proposed a statistical measure of saturation based on the frequency of codes across interviews. These methods differ in detail, but they share a goal: to replace an unsupported claim of saturation with a reported procedure that readers can evaluate.

Information power

Malterud et al. (2016) proposed an alternative way of thinking about sample size, which they called **information power**. The principle is simple: the more information relevant to the study that the sample holds, the fewer participants are needed. They identified five dimensions that affect information power.

Information power

What reduces or increases the number of participants needed

Dimension	Fewer participants	More participants
Study aim	Narrow	Broad
Sample specificity	Dense: highly specific to the aim	Sparse: loosely related
Theory	Applies established theory	No theoretical framework
Quality of dialogue	Strong, rich interviews	Weak, thin interviews
Analysis strategy	In-depth case analysis	Exploratory cross-case analysis

Based on Malterud, Siersma & Guassora (2016) ModQu.com

Figure 2. Five dimensions of information power: the more information the sample holds, the fewer participants are needed. Based on Malterud et al. (2016); table by ModQu.

A study with a narrow aim, participants who are highly specific to that aim, a strong theoretical framework, rich dialogue in the interviews and an in-depth analysis of a few cases needs fewer participants than a study with a broad aim, loosely relevant participants, no theoretical framework, thin interviews and a cross-case analysis. Information power does not produce a number, but it provides a structured way to plan an initial sample size and to judge, during the study, whether more participants are needed.

Sim et al. (2018) examined whether qualitative sample size can be determined before a study begins and concluded that a priori estimates can be useful for planning and funding, but should be treated as provisional, to be revised as the study proceeds.

The critique of saturation

Some researchers argue that saturation is not a meaningful concept for certain approaches. Braun and Clarke (2021) argued that in **reflexive thematic analysis**, themes are constructed by the researcher through interpretation rather than discovered in the data, so there is no fixed set of themes waiting to be saturated. More data could always lead to new interpretations. They recommended that researchers using reflexive approaches abandon saturation as a sample size rationale and instead justify sample size in terms of information power, the richness of the data and the aims of the study.

Low (2019) proposed a more pragmatic definition: theoretical saturation is reached when the researcher's theory adequately explains the phenomenon, and new data do not change the explanation, even if they add detail. Fusch and Ness (2015) argued that saturation depends on data quality as well as quantity, and that a few rich interviews may achieve more than many thin ones.

Morse (2000) identified several factors that affect how many participants are needed: the scope of the study, the nature of the topic, the quality of the data, the study design and the use of what she called **shadowed data**, in which participants describe others' experiences as well as their own. Broad questions, complex topics and poor-quality data all require larger samples.

Saturation across methods

Most research on saturation concerns individual interviews, but the question arises in other methods too. In **focus groups**, the unit of saturation is usually the group rather than the individual, because participants influence one another. The evidence cited above suggests that a small number of groups per type of participant captures most themes, but studies comparing several types need groups for each (Guest, Namey, & McKenna, 2017).

In **ethnography**, saturation is rarely discussed in terms of numbers. Fieldwork continues until the researcher can anticipate what will happen in familiar situations and explain why, and until new observations stop altering the account. The judgement is similar to theoretical saturation, but it applies to settings, events and routines as well as to people.

In studies of **documents and online content**, the volume of available material is often far larger than any researcher can read. Here, saturation is usually judged by sampling texts in batches and tracking new codes, much as with interviews. Because such material is easy to collect, the risk is not too little data but too much, analysed too thinly.

In **case studies**, the relevant question is how many cases are needed, which depends on the replication logic of the design rather than on saturation in the usual sense.

Saturation and diversity

Saturation is reached more quickly in homogeneous samples, because participants share more experiences. This creates a temptation: a study can reach saturation early simply by talking to similar people. Saturation within a narrow group says nothing about experiences outside it. If a study aims to cover a diverse population, saturation should be judged within each main subgroup, or the study should be explicit that its findings reflect the range of experience in the groups it included.

The same applies to the people who agree to take part. Those who volunteer for interviews may share characteristics, such as stronger views or more free time, that make their accounts more similar to each other than to those of people who declined. Saturation among volunteers can therefore overstate how completely a topic has been covered.

Saturation in commercial research

Commercial qualitative research usually works to fixed budgets and timelines, and sample sizes are often set before fieldwork begins. This is not necessarily a problem, provided the fixed number is chosen with care and its limits are acknowledged. A planned sample of 20 interviews, with two segments of ten each, may be adequate for a focused question about a fairly uniform customer group, and inadequate for a broad exploratory question across many segments.

Commercial teams can borrow the logic of stopping rules even within a fixed budget. Analysing interviews as they are completed, rather than all at the end, allows the team to see whether new themes are still appearing and to adjust the remaining interviews towards gaps, for example by recruiting more of a segment whose views are least understood. A brief note in the report on how much new material the final interviews produced gives clients a realistic sense of how complete the findings are.

Practical guidance

Drawing these strands together, a defensible approach to qualitative sample size has several elements.

Plan a range, not a number. Use information power, published evidence on similar studies, and the practical constraints of time and budget to set an initial range. For a focused study of a fairly homogeneous group using interviews, evidence suggests that 12 to 20 interviews is often sufficient to identify the main themes. Studies that compare several groups need enough participants within each group. Baker and Edwards (2012) collected the views of leading qualitative methodologists on the question and found that most answered "it depends", while offering guidance on the factors that matter.

Sample purposively and theoretically. Choose participants who are likely to hold relevant and varied experiences. If rare experiences matter, seek them out deliberately rather than hoping they appear by chance (Small, 2009).

Analyse as you go. Saturation cannot be judged after all data have been collected. Code interviews as they are completed, track when new codes and themes appear, and adjust the sample in response.

Use a stated stopping rule where saturation is the rationale. Define in advance what will count as saturation and how it will be assessed, for example a base of 10 interviews followed by runs of three, stopping when a run adds fewer than 5% new codes.

Distinguish identifying from understanding. Hearing an issue mentioned is not the same as understanding it. Meaning saturation typically requires more data than code saturation (Hennink et al., 2017).

Report honestly. State the sample size rationale, the meaning of saturation used, how it was assessed and any limitations. If the sample was fixed by budget, say so, and discuss what that implies.

Reporting sample size

A short, specific statement of how sample size was determined is one of the simplest ways to strengthen a qualitative report. Vasileiou et al. (2018) recommended that researchers describe their sample size rationale explicitly, rather than relying on a passing reference to saturation. A useful statement covers:

- the initial planned sample and the reasoning behind it, such as information power or published evidence from similar studies;
- the sampling strategy, including any deliberate search for rare or contrasting experiences;
- which meaning of saturation, if any, was used, and why it fits the analytic approach;
- how saturation or sufficiency was assessed, including any stopping rule and the evidence that it was met;
- any constraints, such as budget or access, that limited the sample, and what they imply for the findings.

A table showing when each code or theme first appeared, or how many new codes each batch of interviews produced, makes the claim concrete. Guest et al. (2020) suggested reporting saturation as a simple figure, such as the percentage of new information added by the last run of interviews, so that readers can see how close to saturation the study came.

A worked example

Consider a hypothetical study of why small retailers adopt or reject a new payment system. The team plans 15 to 25 interviews, using information power: the aim is focused, participants are all small retail owners who have considered the system, and interviews are expected to be detailed. They plan to sample both adopters and non-adopters, and to seek out at least three owners who adopted and then abandoned the system, because this experience is rare but important.

They analyse interviews in batches. After the first 12, they have identified 38 codes. The next three interviews add two new codes; the following three add none. By the stopping rule they set in advance, they have reached code saturation at 18 interviews. However, two codes concerning abandonment remain thinly understood, so they conduct three further interviews with owners who abandoned the system, reaching 21. The report describes this process, including the deliberate extension for meaning saturation on one theme.

Conclusion

Saturation is a useful idea but a slippery one. It began as a specific concept in grounded theory, concerning the development of theoretical categories, and has spread to mean several different things. Empirical studies suggest that in focused studies of fairly homogeneous groups, the main themes often appear within about a dozen interviews, but that understanding them fully, and capturing rarer experiences, takes more. The best practice is to plan a sample using principles such as information power, sample purposively, analyse during data collection, use explicit criteria for stopping, and report the process transparently. The question is not only "how many?" but "how will we know we have enough, and how will we show it?".

References

Baker, S. E., & Edwards, R. (2012). *How many qualitative interviews is enough? Expert voices and early career reflections on sampling and cases in qualitative research*. National Centre for Research Methods. https://eprints.ncrm.ac.uk/id/eprint/2273/

Braun, V., & Clarke, V. (2021). To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. *Qualitative Research in Sport, Exercise and Health, 13*(2), 201–216. https://doi.org/10.1080/2159676X.2019.1704846

Charmaz, K. (2014). *Constructing grounded theory* (2nd ed.). SAGE.

Dey, I. (1999). *Grounding grounded theory: Guidelines for qualitative inquiry*. Academic Press.

Francis, J. J., Johnston, M., Robertson, C., Glidewell, L., Entwistle, V., Eccles, M. P., & Grimshaw, J. M. (2010). What is an adequate sample size? Operationalising data saturation for theory-based interview studies. *Psychology & Health, 25*(10), 1229–1245. https://doi.org/10.1080/08870440903194015

Fusch, P. I., & Ness, L. R. (2015). Are we there yet? Data saturation in qualitative research. *The Qualitative Report, 20*(9), 1408–1416. https://doi.org/10.46743/2160-3715/2015.2281

Galvin, R. (2015). How many interviews are enough? Do qualitative interviews in building energy consumption research produce reliable knowledge? *Journal of Building Engineering, 1*, 2–12. https://doi.org/10.1016/j.jobe.2014.12.001

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Aldine.

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods, 18*(1), 59–82. https://doi.org/10.1177/1525822X05279903

Guest, G., Namey, E., & Chen, M. (2020). A simple method to assess and report thematic saturation in qualitative research. *PLOS ONE, 15*(5), Article e0232076. https://doi.org/10.1371/journal.pone.0232076

Guest, G., Namey, E., & McKenna, K. (2017). How many focus groups are enough? Building an evidence base for nonprobability sample sizes. *Field Methods*, *29*(1), 3–22. https://doi.org/10.1177/1525822X16639015

Hennink, M., & Kaiser, B. N. (2022). Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Social Science & Medicine*, *292*, Article 114523. https://doi.org/10.1016/j.socscimed.2021.114523

Hennink, M. M., Kaiser, B. N., & Marconi, V. C. (2017). Code saturation versus meaning saturation: How many interviews are enough? *Qualitative Health Research*, *27*(4), 591–608. https://doi.org/10.1177/1049732316665344

Low, J. (2019). A pragmatic definition of the concept of theoretical saturation. *Sociological Focus*, *52*(2), 131–139. https://doi.org/10.1080/00380237.2018.1544514

Lowe, A., Norris, A. C., Farris, A. J., & Babbage, D. R. (2018). Quantifying thematic saturation in qualitative data analysis. *Field Methods*, *30*(3), 191–207. https://doi.org/10.1177/1525822X17749386

Malterud, K., Siersma, V. D., & Guassora, A. D. (2016). Sample size in qualitative interview studies: Guided by information power. *Qualitative Health Research*, *26*(13), 1753–1760. https://doi.org/10.1177/1049732315617444

Mason, M. (2010). Sample size and saturation in PhD studies using qualitative interviews. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *11*(3), Article 8. https://doi.org/10.17169/fqs-11.3.1428

Morse, J. M. (1995). The significance of saturation. *Qualitative Health Research*, *5*(2), 147–149. https://doi.org/10.1177/104973239500500201

Morse, J. M. (2000). Determining sample size. *Qualitative Health Research*, *10*(1), 3–5. https://doi.org/10.1177/104973200129118183

O'Reilly, M., & Parker, N. (2013). 'Unsatisfactory saturation': A critical exploration of the notion of saturated sample sizes in qualitative research. *Qualitative Research*, *13*(2), 190–197. https://doi.org/10.1177/1468794112446106

Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., Burroughs, H., & Jinks, C. (2018). Saturation in qualitative research: Exploring its conceptualization and operationalization. *Quality & Quantity*, *52*(4), 1893–1907. https://doi.org/10.1007/s11135-017-0574-8

Sim, J., Saunders, B., Waterfield, J., & Kingstone, T. (2018). Can sample size in qualitative research be determined a priori? *International Journal of Social Research Methodology*, *21*(5), 619–634. https://doi.org/10.1080/13645579.2018.1454643

Small, M. L. (2009). "How many cases do I need?" On science and the logic of case selection in field-based research. *Ethnography*, *10*(1), 5–38. https://doi.org/10.1177/1466138108099586

van Rijnsoever, F. J. (2017). (I can't get no) saturation: A simulation and guidelines for sample sizes in qualitative research. *PLOS ONE, 12*(7), Article e0181689. https://doi.org/10.1371/journal.pone.0181689

Vasileiou, K., Barnett, J., Thorpe, S., & Young, T. (2018). Characterising and justifying sample size sufficiency in interview-based studies: Systematic analysis of qualitative health research over a 15-year period. *BMC Medical Research Methodology, 18*, Article 148. https://doi.org/10.1186/s12874-018-0594-7