RESEARCH · QUALITATIVE RESEARCH

Sampling Logic in Qualitative Research and Its Implications for Knowledge Transferability

Qualitative studies select participants for what they can reveal, not to represent a population statistically. This article explains purposeful sampling, how sample size is justified, and how findings from small samples can still travel to other settings.

ModQu research article · 2,938 words · 26 references (APA 7th edition)

Qualitative researchers are often asked a question that comes from survey research: "How can you generalise from twenty interviews?" The question assumes that the purpose of a sample is to estimate a population value, and that larger and more random samples are always better. Qualitative research works on a different logic. It selects people, settings and cases because they can shed light on a process, an experience or a meaning, and it makes claims about how and why things happen rather than how often. Small samples are not a weakness of qualitative research but a consequence of its aims. They do, however, place a heavy responsibility on the researcher to choose cases well, to justify the number chosen and to explain clearly to whom and where the findings may apply.

This article sets out the logic of qualitative sampling, describes the main strategies, discusses how sample size is justified and explains the different ways in which qualitative knowledge can be transferred beyond the people studied.

Two logics of sampling

In **probability sampling**, every member of a population has a known chance of selection, and the sample is used to estimate population values with a measurable margin of error. The logic is *statistical*: the sample stands for the population because chance selected it.

In **purposeful sampling**, cases are selected deliberately because they are "information-rich", able to reveal a great deal about the issues central to the research (Patton, 2015). The logic is *analytical*: the sample is useful because of what the cases can teach, not because they mirror the population in miniature. Marshall (1996) argued that random sampling is often inappropriate for qualitative research, because the aim is to understand complex human issues and the people best able to illuminate them are not evenly spread through a population.

The difference affects every later decision. A survey of 1,000 customers chosen at random can tell a company what proportion are dissatisfied. Twenty in-depth interviews chosen purposefully, including long-standing loyal customers, recent defectors and customers who complained and stayed, can tell the company *why* people become dissatisfied and what keeps some of them loyal. The second study cannot estimate proportions; the first cannot explain mechanisms in depth. Each answers its own question.

Small (2009) made the point sharply for ethnographic and interview studies. Researchers who try to make small qualitative samples look "representative", for example by selecting 35 interviewees in proportion to population groups, gain little: the sample is still too small for statistical inference and the attempt distracts from the real strengths of qualitative design. He proposed that qualitative studies should instead be understood through the logic of case studies and of sequential interviewing, in which each case adds to understanding and later cases are chosen to test and refine what earlier ones suggested.

Purposeful sampling strategies

Patton (2015) described many purposeful sampling strategies. Figure 1 groups the most common ones by their purpose.

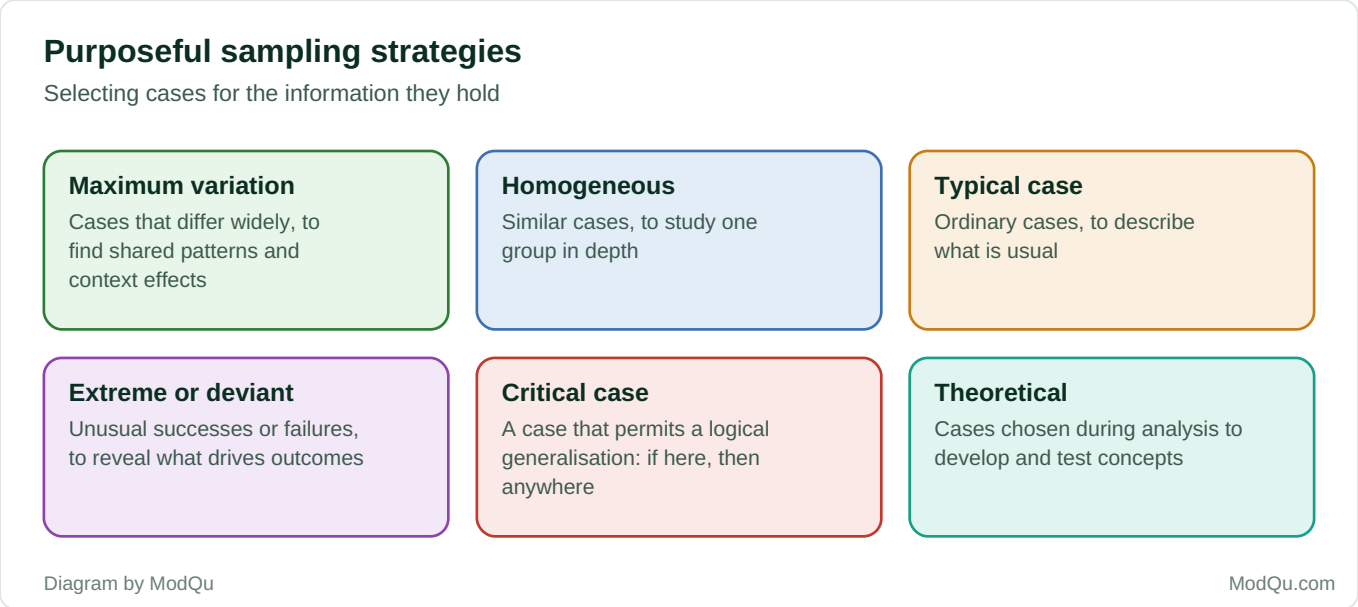


Figure 1. Common purposeful sampling strategies and what each is for, based on Patton (2015). Diagram by ModQu.

Maximum variation sampling selects cases that differ widely on dimensions thought to matter. Patterns that appear across very different cases are likely to be important and shared, while differences reveal how context shapes experience.

Homogeneous sampling selects similar cases to study one group in depth, as is common in focus groups.

Typical case sampling selects cases that are average or normal, to describe what is usual.

Extreme or deviant case sampling selects unusual cases, such as outstanding successes or notable failures, because they can reveal the conditions that produce exceptional outcomes.

Critical case sampling selects a case that allows a logical generalisation: "if it happens here, it will happen anywhere", or "if it does not happen here, it will not happen anywhere".

Theoretical sampling, developed in grounded theory, selects cases as the analysis proceeds, to develop and test emerging concepts (Glaser & Strauss, 1967). Charmaz (2014) emphasised that theoretical sampling aims to develop the properties of categories, not to represent a population.

Other strategies include snowball sampling, in which participants refer others, which is useful for hidden or hard-to-reach groups but tends to recruit people who are connected to one another (Faugier & Sargeant, 1997); and convenience sampling, which is the weakest because it is driven by access rather than information.

Robinson (2014) proposed a four-point approach for interview studies: define the sample universe through inclusion and exclusion criteria; decide the sample size; choose a sampling strategy; and plan how to source participants. Making each decision explicit helps readers judge what the findings can support. Palinkas et al. (2015) reviewed the use of purposeful sampling in mixed methods implementation research and noted that studies often combined strategies, for example selecting typical and extreme cases from a larger quantitative sample, and that the choice should follow the study's purpose.

Case selection

When the unit of study is an organisation, a programme or a community rather than a person, sampling becomes **case selection**. Yin (2018) distinguished single-case designs, appropriate for critical, unusual, common, revelatory or longitudinal cases, from multiple-case designs, in which cases are chosen to replicate one another. Literal replication selects cases expected to produce similar results; theoretical replication selects cases expected to produce different results for predictable reasons.

Seawright and Gerring (2008) described seven techniques for selecting cases, including typical, diverse, extreme, deviant, influential, most-similar and most-different cases, and showed how each can be chosen systematically from a larger data set when one exists. Stake (1995) distinguished intrinsic case studies, chosen because the case itself is of interest, from instrumental case studies, chosen because they help understand a broader issue. The kind of claim a researcher can make depends on which of these choices was made.

How many is enough?

Qualitative sample size has generated a large literature, because reviewers, ethics committees and funders want a number in advance while the logic of qualitative inquiry suggests the number should depend on what the data reveal.

Saturation

The most common justification is **saturation**, the point at which additional data no longer produce new information relevant to the research question. The idea comes from grounded theory, where theoretical saturation meant that no new properties of a category were emerging (Glaser & Strauss, 1967).

Guest et al. (2006) examined saturation empirically. In a study of 60 interviews with women in two West African countries, they recorded when each code first appeared and found that most codes emerged within the first twelve interviews, with the basic elements of the main themes present within the first six. Hennink et al. (2017) distinguished

code saturation, the point at which no new issues are identified, from **meaning saturation**, the point at which issues are fully understood. In their study, code saturation was reached at about nine interviews, but meaning saturation needed considerably more.

Figure 2 shows why these numbers vary. In a simple model, each theme appears in an interview with a certain probability. Common themes are found quickly; rare themes take many interviews. A homogeneous sample with mostly common themes reaches saturation quickly. A heterogeneous sample, in which many themes are specific to particular subgroups, takes much longer.

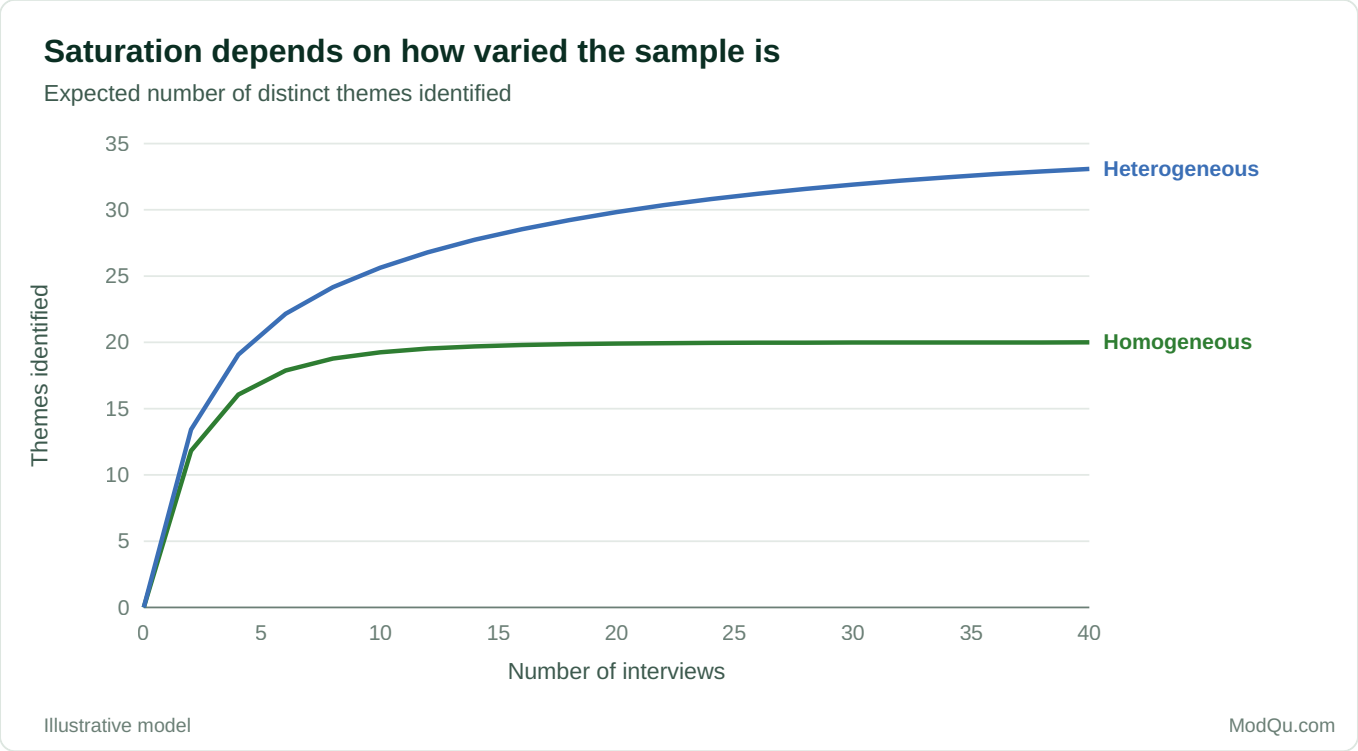


Figure 2. Expected themes identified as interviews are added. Each theme appears in an interview with a fixed probability; the homogeneous sample has 20 themes with probabilities 0.15 to 0.6, and the heterogeneous sample adds 15 themes with probabilities 0.03 to 0.08. Illustrative model; chart by ModQu.

Saturation has been criticised. Saunders et al. (2018) identified several different meanings of saturation in the literature, some concerning theory, some data and some codes, and warned that the term is often used without saying which. Braun and Clarke (2021) argued that saturation fits poorly with reflexive thematic analysis, in which themes are constructed by the researcher rather than discovered in the data, so there is no fixed set of themes waiting to be found. They recommended that researchers justify sample size by reference to the richness of the data and the aims of the analysis.

Information power

Malterud et al. (2016) proposed **information power** as an alternative. The more relevant information a sample holds, the fewer participants are needed. Information power depends on five factors:

1. **Aim:** a narrow aim needs fewer participants than a broad one.
2. **Specificity:** participants with experiences highly specific to the aim provide more information.
3. **Theory:** studies applying established theory need fewer participants than purely exploratory ones.
4. **Dialogue quality:** strong, focused interviews yield more than weak ones.
5. **Analysis strategy:** in-depth analysis of a few cases needs fewer than cross-case analysis of many.

Information power can be assessed at the planning stage and reviewed during data collection, which makes it practical for proposals and ethics applications.

Practice and pitfalls

Mason (2010) reviewed doctoral theses using qualitative interviews and found that sample sizes clustered at round numbers, such as 20 and 30, suggesting that sizes were often chosen by convention rather than by any principle. Morse (2000) argued that the number needed depends on the scope of the study, the nature of the topic, the quality of the data and the design, and that a study with a clear focus and articulate participants needs fewer people than one with a broad, diffuse question.

Small samples also bring practical risks. If key subgroups are missing, saturation may be declared within a narrow slice of experience. Researchers should therefore define the dimensions along which experience is expected to vary and check that the sample covers them, rather than stopping when interviews with similar people stop producing surprises.

Sampling in commercial qualitative research

In market and business research, qualitative samples are usually recruited through screening questionnaires that check whether a person meets the inclusion criteria. Screeners translate the sampling strategy into practice, so their design matters. A screener that asks only about product use will produce a sample of users; one that also asks about previous brands, frequency, attitudes and circumstances allows maximum variation to be built in deliberately.

Three risks are common. The first is **professional respondents**, people who take part in many studies and learn what researchers want to hear; recruiters often exclude anyone who has taken part in research recently. The second is **recruitment by convenience**, in which the easiest people to reach, such as the most engaged customers on a mailing list, dominate the sample. The third is **quota thinking**, in which quotas copied from a survey are applied to a sample of twelve, giving one or two people per cell and an appearance of representativeness without its substance. Small (2009) warned against exactly this. Quotas in qualitative work are better used to guarantee coverage of the dimensions that matter than to mirror population shares.

Remote and online methods have widened the pool of possible participants, making it easier to include people from different regions and those who cannot travel. They can also exclude people with limited internet access or low confidence with technology, so the sampling plan should consider who the mode leaves out.

Transferability and other forms of generalisation

If qualitative findings are not statistically generalisable, how can they be useful beyond the people studied? Several answers have been proposed, and they are complementary rather than competing. Figure 3 summarises three main routes.

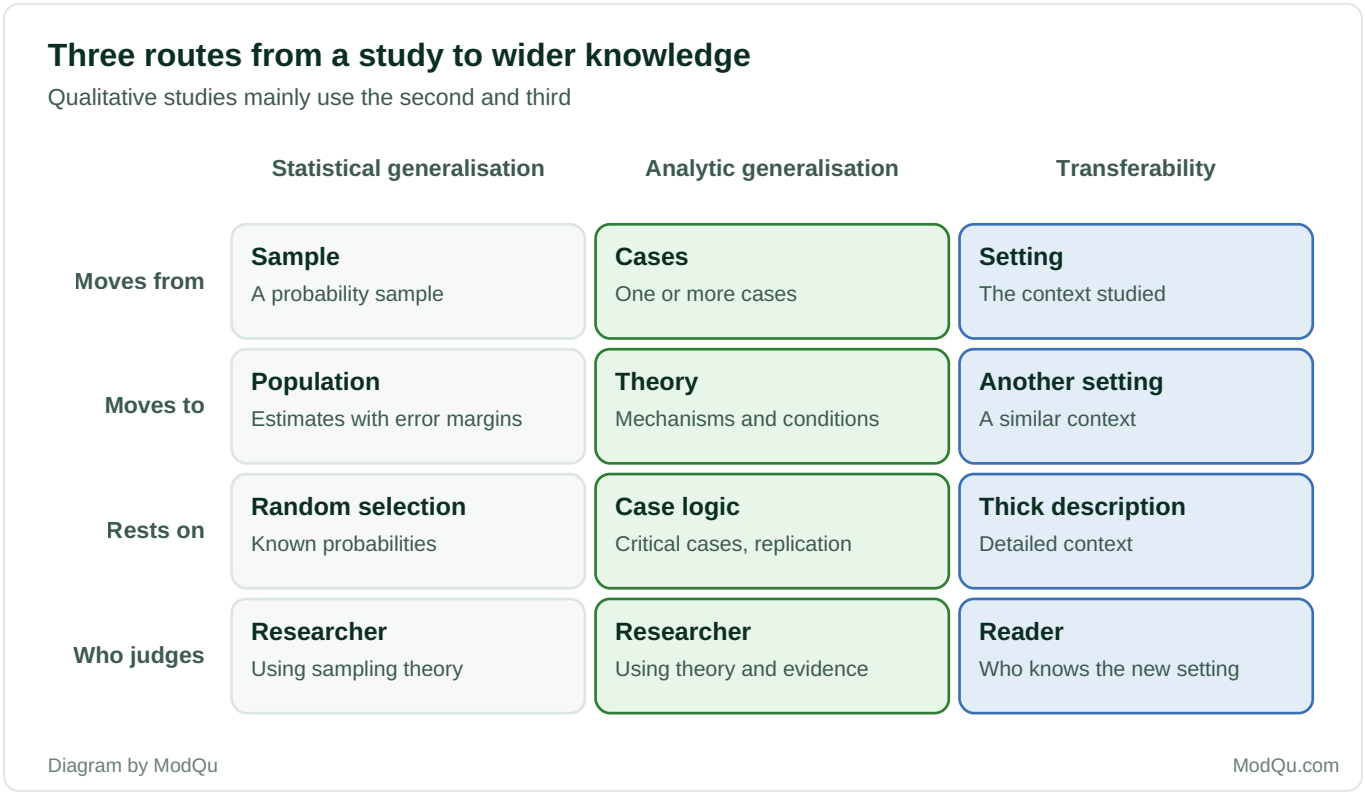


Figure 3. Three routes by which findings travel beyond the study. Based on Firestone (1993), Lincoln and Guba (1985) and Yin (2018). Diagram by ModQu.

Analytic generalisation

Firestone (1993) distinguished three arguments for generalising: from sample to population, analytic generalisation and case-to-case transfer. In **analytic generalisation**, findings are generalised to a theory rather than a population. A case study does not show how often a mechanism operates, but it can show that it operates, how it works and under what conditions. Yin (2018) argued that case studies generalise to theoretical propositions in the same way that experiments do: a single laboratory experiment is not representative of anything, but its results are used to support or revise a theory.

Flyvbjerg (2006) challenged the view that one cannot generalise from a single case. He argued that a well-chosen critical case can falsify a general proposition, that much important knowledge in many fields comes from detailed cases, and that context-dependent knowledge is the basis of expertise.

Transferability

Lincoln and Guba (1985) proposed **transferability** as the qualitative counterpart of external validity. The researcher cannot know where findings will be applied, so the responsibility for judging whether they fit another setting lies with the person who wishes to apply them. The researcher's responsibility is to provide enough detail about the setting, participants and processes, often called **thick description**, for that judgement to be made. Transferability is one element of the broader idea of **trustworthiness**.

Polit and Beck (2010) suggested practical strategies to strengthen transferability: describe participants and contexts in detail, sample for variation, look for confirming and disconfirming cases, compare findings with other studies and be explicit about the conditions under which findings may not hold. Tracy (2010) included "resonance", the ability of research to affect readers and transfer across contexts, among her eight criteria for excellent qualitative research.

Moderatum generalisation

Williams (2000) argued that interpretive researchers routinely make modest generalisations, which he called *moderatum* generalisations: claims that aspects of a situation are instances of a broader set of recognisable features. These are less ambitious than statistical generalisations but more than purely local descriptions, and they are open to testing by further research.

Smith (2018) brought these ideas together and added further forms, such as **naturalistic generalisation**, in which readers recognise similarities between the research and their own experience, and **intersectional generalisation**, in which findings are connected to findings from other studies. He argued that qualitative researchers should state which kind of generalisation they are claiming rather than either overclaiming or apologising.

Validity and the inferential link

Maxwell (1992) distinguished internal generalisation, within the setting or group studied, from external generalisation, beyond it. Even within a study, a researcher who interviews ten staff members in an organisation of five hundred needs to justify claims about "the organisation". Purposeful sampling, including deliberate inclusion of people likely to see things differently, makes internal generalisation more defensible.

A worked example

Suppose a hypothetical software company wants to understand why small businesses stop using its accounting product within the first three months. Its customer database records cancellations but not reasons.

A probability survey of all cancelling customers could estimate how often each listed reason is given. But the company does not yet know what the reasons are, and the reasons customers tick may not be the real ones. The research team therefore plans a qualitative study first.

They define the sample universe as owners or managers of businesses with fewer than 50 staff who cancelled within 90 days in the past six months. Using **maximum variation sampling**, they select interviewees across industries, business sizes, the customer's previous accounting method and whether the customer had contacted support. They also add a few **deviant cases**: similar businesses that nearly cancelled but stayed after contacting support.

They plan about 20 interviews, justified by information power: the aim is narrow, participants have highly specific experience, interviews are focused and the analysis is thematic. After 14 interviews, new codes are rare among cancelling customers, but the deviant cases continue to produce new insights about what support did differently. The team conducts six more interviews with such customers.

The findings identify a mechanism: customers who imported data from spreadsheets encountered errors that support could fix quickly, but most never contacted support and left. The team does not claim to know what proportion of cancellations this explains. That is a question for the next stage, a survey of cancelling customers with response options informed by the interviews. The qualitative findings transfer analytically, to the theory of early-stage abandonment, and practically, to other products with similar onboarding steps, with thick description that allows other teams to judge fit.

Reporting sampling decisions

Readers cannot judge transferability without knowing how the sample was built. A qualitative report should state the sample universe, the strategy or strategies used, how participants were identified and recruited, how many were approached and how many took part, the characteristics of those who took part, how sample size was decided and when recruitment stopped, and which perspectives may be missing. Reporting these details is not a ritual; it is the evidence on which any claim to wider relevance rests.

Practical guidance

1. **State the purpose of sampling.** Explain what each sampling decision was meant to achieve.
2. **Define the sample universe.** Give clear inclusion and exclusion criteria.
3. **Choose strategies deliberately.** Use maximum variation, deviant cases or theoretical sampling where they serve the question, and combine them if needed.
4. **Justify sample size.** Use information power or a clearly defined form of saturation, and report how and when the decision to stop was made.
5. **Look for disconfirming cases.** Seek people and settings likely to challenge emerging findings.
6. **Describe richly.** Provide enough detail about participants and contexts for readers to judge transferability.
7. **Claim the right kind of generalisation.** Say whether findings generalise to theory, transfer to similar contexts or describe only the setting studied.
8. **Sequence methods.** Where population estimates are needed, follow qualitative work with a survey designed using its findings.

Key takeaways

Qualitative research samples for information, not for statistical representation. Purposeful sampling strategies select cases for what they can reveal, and sample size is justified by the information the sample holds, not by formulas designed for estimation. Qualitative findings travel beyond the study through analytic generalisation to theory, through transferability judged by readers with the help of thick description, and through modest generalisations that further research can test. When researchers state clearly how they sampled and what kind of generalisation they claim, small samples produce knowledge that is both rigorous and useful.

References

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.

Faugier, J., & Sargeant, M. (1997). Sampling hard to reach populations. *Journal of Advanced Nursing, 26*(4), 790–797. https://doi.org/10.1046/j.1365-2648.1997.00371.x

Firestone, W. A. (1993). Alternative arguments for generalizing from data as applied to qualitative research. *Educational Researcher, 22*(4), 16–23. https://doi.org/10.3102/0013189X022004016

Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative Inquiry, 12*(2), 219–245. https://doi.org/10.1177/1077800405284363

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

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

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE.

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

Marshall, M. N. (1996). Sampling for qualitative research. *Family Practice, 13*(6), 522–525. https://doi.org/10.1093/fampra/13.6.522

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

Maxwell, J. A. (1992). Understanding and validity in qualitative research. *Harvard Educational Review, 62*(3), 279–300. https://doi.org/10.17763/haer.62.3.8323320856251826

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

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research, 42*(5), 533–544. https://doi.org/10.1007/s10488-013-0528-y

Patton, M. Q. (2015). *Qualitative research & evaluation methods: Integrating theory and practice* (4th ed.). SAGE.

Polit, D. F., & Beck, C. T. (2010). Generalization in quantitative and qualitative research: Myths and strategies. *International Journal of Nursing Studies, 47*(11), 1451–1458. https://doi.org/10.1016/j.ijnurstu.2010.06.004

Robinson, O. C. (2014). Sampling in interview-based qualitative research: A theoretical and practical guide. *Qualitative Research in Psychology, 11*(1), 25–41. https://doi.org/10.1080/14780887.2013.801543

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

Seawright, J., & Gerring, J. (2008). Case selection techniques in case study research: A menu of qualitative and quantitative options. *Political Research Quarterly, 61*(2), 294–308. https://doi.org/10.1177/1065912907313077

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

Smith, B. (2018). Generalizability in qualitative research: Misunderstandings, opportunities and recommendations for the sport and exercise sciences. *Qualitative Research in Sport, Exercise and Health, 10*(1), 137–149. https://doi.org/10.1080/2159676X.2017.1393221

Stake, R. E. (1995). *The art of case study research*. SAGE.

Tracy, S. J. (2010). Qualitative quality: Eight "big-tent" criteria for excellent qualitative research. *Qualitative Inquiry, 16*(10), 837–851. https://doi.org/10.1177/1077800410383121

Williams, M. (2000). Interpretivism and generalisation. *Sociology*, *34*(2), 209–224. https://doi.org/10.1177/S0038038500000146

Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). SAGE.