RESEARCH · RESEARCH DESIGN

Case Selection Strategies in Multiple-Case Research Designs

How to choose cases in a multiple-case study, covering replication logic, the main selection techniques from typical to most different, negative cases, selection bias, quantitative tools for choosing cases, how many cases to study, and how the same reasoning applies to commercial research.

ModQu research article · 2,466 words · 23 references (APA 7th edition)

A **multiple-case study** examines two or more cases in depth, such as several firms, schools, hospitals or product launches. Because each case takes considerable time and money, most studies of this kind include between two and a dozen cases. With so few, the choice of which ones to study shapes almost everything the study can conclude. A survey can rely on random sampling to make its sample resemble the population. A case study usually cannot. The researcher must choose cases deliberately, and the reasoning behind that choice is part of the method, not a preliminary to it. This article sets out the main strategies for selecting cases, the logic behind each, and the mistakes that most often weaken case-based conclusions.

Why case selection is different

Gerring (2004) defined a case study as the intensive study of a single unit for the purpose of understanding a larger class of similar units. A multiple-case study extends this to several units, but the aim is the same: to learn something that reaches beyond the cases themselves. That aim makes selection difficult. If cases are picked because they are convenient, well known or successful, the conclusions may say more about the selection than about the class of cases.

Small (2009) argued that researchers using case and interview methods often borrow the language of survey sampling, asking whether their cases are "representative", when their studies cannot meet the conditions that make that language meaningful. With five cases, even random selection can easily produce an unusual set. Seawright and Gerring (2008) made the same point: random sampling works well for large samples, but for very small samples it often yields cases that are unrepresentative or uninformative, so purposive selection is usually the better choice. The question for the case researcher is therefore not "how do I get a representative sample?" but "which cases will be most informative for the question I am asking?"

Replication logic

Yin (2018) offered the most widely used answer for multiple-case designs. Each case is treated not as a sampling unit but as a separate experiment. Just as a scientist repeats an experiment to see whether the result holds, the case researcher studies further cases to see whether the pattern found in the first one recurs. Yin distinguished two kinds of replication:

- **Literal replication:** cases are chosen because they are expected to produce similar results, which tests whether a finding holds across similar conditions.
- **Theoretical replication:** cases are chosen because the theory predicts contrasting results for predictable reasons, which tests whether the theory explains the differences.

Eisenhardt (1989) took a similar view in her account of building theory from cases. Cases are chosen for **theoretical** reasons, to replicate or extend emerging theory, to fill conceptual categories or to provide **polar types**, rather than for statistical ones. Eisenhardt and Graebner (2007) later answered critics who saw this as a weakness, arguing that theoretical sampling is appropriate precisely because the goal is to develop theory, not to estimate how common something is in a population. The approach draws on **theoretical sampling** from grounded theory, in which each new case or informant is chosen to develop concepts that the analysis so far has raised (Glaser & Strauss, 1967).

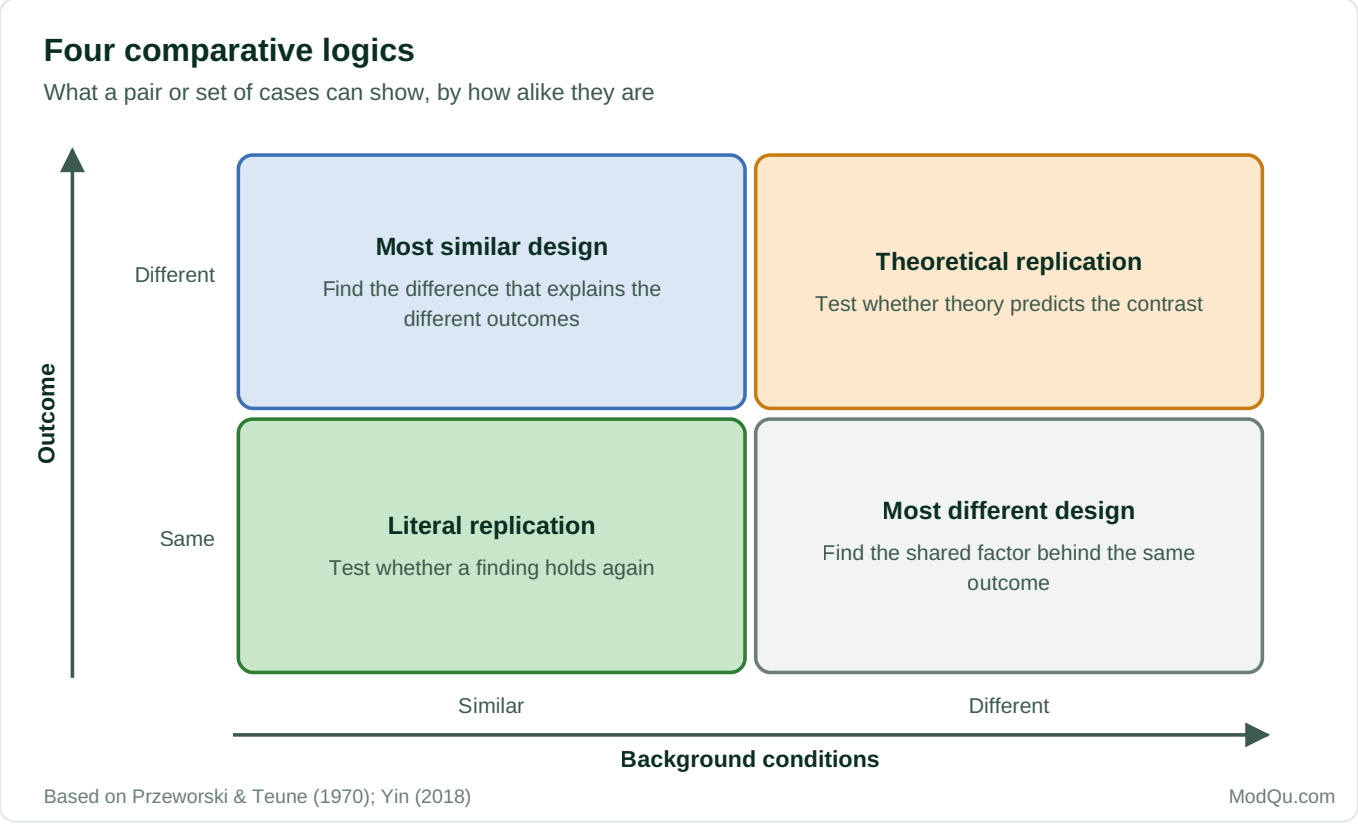


Figure 1. Four ways to combine cases, by how alike their background conditions and outcomes are. Based on Przeworski and Teune (1970) and Yin (2018); diagram by ModQu.

The matrix above shows how these logics combine with the classic comparative designs. Choosing cases that are alike in their background conditions but differ in outcome isolates the factor that might explain the difference. Choosing cases that differ in many respects but share an outcome looks for the factor they have in common.

Seven techniques for choosing cases

Seawright and Gerring (2008) set out a menu of seven techniques, each suited to a different purpose. The table below summarises them.

Seven ways to choose cases

Each technique answers a different kind of question

Technique	How cases are chosen	Used to
Typical	Fit a known pattern well	Show the mechanism behind it
Diverse	Cover the full range of values	Capture variety in the population
Extreme	Unusually high or low on a cause or outcome	Explore a phenomenon in clear form
Deviant	Do not fit the expected pattern	Find factors theory has missed
Influential	Change a statistical result if removed	Check a model's key assumptions
Most similar	Alike except on the cause and outcome	Isolate a causal factor
Most different	Unlike except on the cause and outcome	Find a shared causal factor

Based on Seawright & Gerring (2008) ModQu.com

Figure 2. Seven techniques for selecting cases and the purpose each serves. Based on Seawright and Gerring (2008); table by ModQu.

Typical cases are those that fit a known pattern well. They are useful for probing the mechanism behind a relationship that is already established statistically: if larger firms are more likely to adopt a practice, a typical large adopter can show how and why. **Diverse** cases cover the full range of values on a key variable or the full set of combinations of several variables, so that the study captures the variety in the population.

Extreme cases have unusually high or low values on a cause or outcome. They are often used in exploratory work, because an extreme case shows a phenomenon in its clearest form. **Deviant** cases are those that do not fit the expected pattern: a firm that should have failed by every known predictor but thrived, or a region with all the conditions for growth that stagnated. Deviant cases are valuable for finding factors that existing theory has missed.

Influential cases are those whose inclusion or exclusion changes a statistical result substantially. Examining them in depth can show whether they are measurement errors, special circumstances or evidence that the model is wrong. **Most similar** and **most different** cases follow the comparative designs described by Przeworski and Teune (1970), shown in the earlier matrix.

These techniques are not exhaustive. Eckstein (1975) argued for the **crucial case**: one in which a theory is most likely, or least likely, to hold. If a theory fails where it should most clearly succeed, or succeeds where it should most clearly fail, the result is strong evidence. Flyvbjerg (2006) developed a related idea, the **critical case**, which allows the logic "if it is valid for this case, it is valid for all (or many) cases". A study of whether a new safety procedure is followed, for example, might deliberately choose the site where conditions for compliance are best: if staff there do not follow it, few others will.

Purposive selection in qualitative research

Qualitative researchers have developed a parallel vocabulary. Patton (2015) described many forms of **purposeful sampling**, including intensity sampling (information-rich cases that show the phenomenon strongly but not unusually), maximum variation sampling (cases that differ widely, to identify shared patterns that cut across variation), and homogeneous sampling (similar cases, to describe one subgroup in depth). These map closely onto the techniques above: maximum variation corresponds to diverse cases, intensity to a moderated form of extreme cases.

Stake (2006) emphasised that in multiple-case research the object of interest is the **quintain**, the program, phenomenon or condition that the cases share. He suggested three criteria for choosing cases: relevance to the quintain, diversity across contexts, and good opportunities to learn about complexity and context. The last point is practical. A theoretically ideal case is of little use if access is poor, records are missing or key people will not speak.

Selection bias

The most serious risk in case selection is choosing cases on the basis of the outcome. Geddes (1990) showed how studies that examine only cases where an outcome occurred, such as only countries that experienced rapid growth, can reach conclusions that disappear when cases without the outcome are included. A factor found in every successful case may be just as common among the failures. King et al. (1994) gave the general rule: selecting on the dependent variable restricts its variation and biases estimates of causal effects, typically towards zero.

Collier and Mahoney (1996) argued that the problem is real but more nuanced in qualitative research. Selecting on the outcome is acceptable, and sometimes necessary, when the aim is to identify necessary conditions or to trace how an outcome came about in specific cases, rather than to estimate the average effect of a cause. The danger lies in drawing general causal conclusions from a set of cases that includes only one kind of outcome.

Negative cases and the possibility principle

Including cases where the outcome did not occur solves one problem but raises another: which non-occurrences? There are countless cases in which, say, a revolution or a market breakthrough did not happen. Mahoney and Goertz (2004) proposed the **possibility principle**: negative cases should be those where the outcome was possible, meaning that at least one of the conditions theory associates with the outcome was present. A study of why some firms successfully entered a foreign market should compare them with firms that tried, or were well placed to try, and failed, not with firms that never had the means or the intention to enter.

Using data to choose cases

Case selection is often stronger when it draws on data about the wider population. Lieberman (2005) proposed **nested analysis**, in which a large-sample statistical analysis comes first and its results guide the choice of cases. If the model fits well, cases on or near the regression line are studied to confirm the mechanism; if it fits poorly, cases off the line are studied to find what the model has missed. The cases and the statistics then inform each other.

Several later studies tested these ideas formally. Herron and Quinn (2016) used simulations to compare how well different selection techniques help a researcher learn about a causal effect, and found substantial differences in performance between them, which suggests that the choice of technique should be justified rather than assumed. Nielsen (2016) showed how statistical **matching**, normally used to estimate treatment effects, can be used to find pairs of cases that are as similar as possible except for the factor of interest, putting the most similar design on a more systematic footing. Gerring and Cojocaru (2016) argued that the right technique depends on the goal of the study, distinguishing descriptive goals, exploratory causal goals, and goals of estimating or diagnosing a known causal relationship, each of which favours different kinds of cases.

How many cases?

There is no fixed number. Eisenhardt (1989) suggested that between four and ten cases usually works well for building theory: with fewer than four, it is hard to generate theory of any complexity, while with more than ten the volume of data becomes difficult to manage. Yin (2018) argued that the number should follow from the replication logic: two or three literal replications may be enough when the theory is straightforward and the differences sought are large, while more are needed when the theory is subtle or rival explanations must be ruled out.

Small (2009) suggested a **sequential** approach in which cases are added one at a time, each chosen in light of what the previous ones revealed, until new cases stop changing the emerging explanation. This mirrors theoretical sampling in grounded theory and is often more efficient than fixing the number in advance. Its drawback is that it makes budgeting harder, so many applied studies set a planned number with an allowance for one or two additional cases.

Practical constraints

In practice, the ideal set of cases is rarely available. Organisations decline to take part, key informants leave, and records turn out to be incomplete. Stake (2006) treated the opportunity to learn as a legitimate selection criterion for this reason. The practical response is to define the selection logic first and then list several candidate cases for each slot in the design. If the preferred deviant case refuses access, a second deviant case can take its place without changing the logic of the study.

Access also introduces its own bias. Organisations that agree to be studied are often those that are proud of their practices or have the staff time to spare, which can tilt a study towards success stories. Recording which cases were approached, which declined and why allows readers to judge how far this has affected the results. Where possible, background data on cases that declined, such as size, sector or performance, can show whether they differ systematically from those that took part.

Timing matters as well. A case chosen as typical at the start of a study may become unusual if its circumstances change during fieldwork, for example after a merger or a change of leadership. Such changes are not necessarily a problem, and may even turn a typical case into an informative deviant one, but the original rationale and the change should both be reported.

Comparing cases systematically

Selection and analysis are linked. George and Bennett (2005) described the method of **structured, focused comparison**: the researcher asks the same set of questions of every case, focused on the aspects relevant to the research objective, so that the findings can be compared systematically. Ragin (1987) argued that comparative researchers should treat each case as a configuration of conditions, because the effect of one factor often depends on the presence of others. Both approaches require cases chosen so that the relevant conditions vary in informative ways, which is another reason to plan selection around the analysis rather than around access alone.

Gibbert et al. (2008) reviewed case studies in leading management journals and found that many reported little about how and why cases were chosen. A study that explains its selection logic allows readers to judge which conclusions the cases can support and which they cannot.

Case selection in commercial research

The same reasoning applies outside academic research. Consider a hypothetical retailer that wants to understand why a new self-service format succeeded in some stores but not others. The example is illustrative. Studying only the five best-performing stores would repeat the error Geddes (1990) described: the features they share may be just as common in stores where the format failed. A stronger design would choose pairs of stores that are similar in size, location type and customer profile but differ in outcome, following the most similar logic, and add one or two deviant cases, such as a store that succeeded despite poor conditions.

Data already held by the business can guide the choice, as nested analysis recommends. A simple model of store performance can identify typical stores, stores that performed far above or below prediction, and stores that are close matches on background conditions. The case work then explains what the model cannot, and the selection logic can be set out in a few sentences in the final report.

Conclusion

In multiple-case research, the cases are not a sample in the statistical sense, so the usual protections of random sampling do not apply. The researcher must choose cases deliberately, for reasons that follow from the research question: to replicate a finding, to contrast conditions, to explore extremes, to explain deviations or to test a theory where it is most or least likely to hold. The main dangers are selecting on the outcome, choosing cases for convenience and failing to explain the choice. A clear selection logic, stated in advance and reported in full, is what allows a handful of cases to support conclusions that reach beyond them.

References

Collier, D., & Mahoney, J. (1996). Insights and pitfalls: Selection bias in qualitative research. *World Politics, 49*(1), 56–91. https://doi.org/10.1353/wp.1996.0023

Eckstein, H. (1975). Case study and theory in political science. In F. I. Greenstein & N. W. Polsby (Eds.), *Handbook of political science: Vol. 7. Strategies of inquiry* (pp. 79–137). Addison-Wesley.

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review, 14*(4), 532–550. https://doi.org/10.5465/amr.1989.4308385

Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. *Academy of Management Journal, 50*(1), 25–32. https://doi.org/10.5465/amj.2007.24160888

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

Geddes, B. (1990). How the cases you choose affect the answers you get: Selection bias in comparative politics. *Political Analysis, 2*, 131–150. https://doi.org/10.1093/pan/2.1.131

George, A. L., & Bennett, A. (2005). *Case studies and theory development in the social sciences*. MIT Press.

Gerring, J. (2004). What is a case study and what is it good for? *American Political Science Review, 98*(2), 341–354. https://doi.org/10.1017/S0003055404001182

Gerring, J., & Cojocaru, L. (2016). Selecting cases for intensive analysis: A diversity of goals and methods. *Sociological Methods & Research, 45*(3), 392–423. https://doi.org/10.1177/0049124116631692

Gibbert, M., Ruigrok, W., & Wicki, B. (2008). What passes as a rigorous case study? *Strategic Management Journal, 29*(13), 1465–1474. https://doi.org/10.1002/smj.722

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

Herron, M. C., & Quinn, K. M. (2016). A careful look at modern case selection methods. *Sociological Methods & Research, 45*(3), 458–492. https://doi.org/10.1177/0049124114547053

King, G., Keohane, R. O., & Verba, S. (1994). *Designing social inquiry: Scientific inference in qualitative research*. Princeton University Press.

Lieberman, E. S. (2005). Nested analysis as a mixed-method strategy for comparative research. *American Political Science Review, 99*(3), 435–452. https://doi.org/10.1017/S0003055405051762

Mahoney, J., & Goertz, G. (2004). The possibility principle: Choosing negative cases in comparative research. *American Political Science Review, 98*(4), 653–669. https://doi.org/10.1017/S0003055404041401

Nielsen, R. A. (2016). Case selection via matching. *Sociological Methods & Research, 45*(3), 569–597. https://doi.org/10.1177/0049124114547054

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

Przeworski, A., & Teune, H. (1970). *The logic of comparative social inquiry*. Wiley-Interscience.

Ragin, C. C. (1987). *The comparative method: Moving beyond qualitative and quantitative strategies*. University of California Press.

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

Stake, R. E. (2006). *Multiple case study analysis*. Guilford Press.

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