# 4 Drawing Contingent Generalizations from Case Studies

Andrew Bennett

## 4.1 Introduction

What lessons can be learned from the international community's slow and piecemeal response to the Ebola epidemic in Guinea, Sierra Leone, and Liberia in 2014? Are the histories and outcomes of microfinance programs in one country or by one lender relevant beyond each country or lender? How can we judge whether the early results of a medical or other experiment are so powerfully indicative of either success or failure that the experiment should be stopped even before all cases are treated or all the evidence is in?

Case studies are one approach to addressing such questions. Yet one of the most common critiques of case study methods is that the results of individual case studies cannot be readily generalized. Oxford professor Bent Flyvbjerg notes that when he first became interested in in-depth case study research in the 1990s, his teachers and colleagues tried to dissuade him from using case studies, arguing "you cannot generalize from a single case." Flyvbjerg concluded that this view constitutes a conventional wisdom that "if not directly wrong, is so oversimplified as to be grossly misleading" (Flyvbjerg, 2006: 219). Similarly, the present chapter notes that the conventional wisdom is not fully wrong, as techniques for generalizing from individual case studies are complex and potentially fallible. The chapter concurs with Flyvbjerg, however, in concluding that we have means of assessing which findings will and will not generalize. For some case studies and some findings, generalization beyond the individual case is not warranted. In other contexts, we can make contingent generalizations from one or more case studies, or

https://doi.org/10.1017/9781108688253.005 Published online by Cambridge University Press

generalizations to a subset of a population that shares a well-defined set of features. In still other instances, sweeping generalizations to large and diverse populations are possible even from a single case study. The answer to whether case studies generalize is "It depends." It depends on our prior causal knowledge, our prior knowledge of populations of cases and of the frequency of contextual variables that enable or disable causal mechanisms, the evidence that emerges from process tracing on case studies (see Chapter 7), and how that evidence updates our prior knowledge of causal mechanisms and the contexts in which they do and do not operate.

A second, and related, critique of case studies is that their findings do not cumulate into successive improvements in theories. The present chapter, in contrast, argues that case studies can contribute to developing two different kinds of progressively better theories. First, case studies can lead to improved theories about individual causal mechanisms and the scope conditions under which they operate. Claims about causal mechanisms are one of the most common kinds of theory in both the social and physical sciences. Second, case studies can contribute to improved "typological theories," or theories about how combinations of causal mechanisms interact in specified issue areas and distributions of resources, stakeholder interests, legitimacy, and institutions. Later case studies can build upon, test, qualify, and extend typological theories developed in earlier ones.

This chapter first clarifies different conceptions of "generalization" in statistical and case study research. It then discusses four kinds of generalization from case studies: generalization from the selection and study of "typical" cases, generalization from most- and least-likely cases, mechanism-based generalization, and generalization via typological theories. The chapter uses studies of the 2014 Ebola epidemic as a running example to illustrate many of these kinds of generalization, and it draws on studies of microfinance programs and medical experiments to illustrate particular kinds of generalization.

## 4.2 Statistical Versus Case Study Views on "Generalization"

While the accurate explanation of individual historical cases is important and useful, the ability to generalize beyond individual cases is rightly considered a key component of both theoretical progress and policy relevance. Theories are abstractions that simplify the task of perceiving and operating in the world, and without some degree and kind of generalization little

https://doi.org/10.1017/9781108688253.005 Published online by Cambridge University Press

simplification is possible. But "generalization" can take on several meanings, and scholars and policy-makers vary in their views on what kinds of generalizations are either possible or pragmatically useful, partly depending on whether their methodological training was mostly in quantitative or qualitative approaches. Thus, it is important to clarify the different meanings that scholars in different methodological traditions typically give to the term "generalization."

Among researchers whose main methods are statistical analysis of observational data, "generalization" is commonly treated as a question of the "average effect" observed between a specified independent variable and the dependent variable of interest in a population. This average effect is represented by the coefficients on the statistically significant independent variables in a regression equation.[1] Similarly, for researchers who use experimental methods, generalization takes the form of the estimated "average treatment effect," measured as the average difference in outcomes between the treated and untreated groups from a large number of randomly selected units.[2]

Generalization from statistical analysis of observational data depends on several assumptions, most notably: 1) that the treatment of one unit does not affect the outcome of another unit (the Stable Unit Treatment Value Assumption, or SUTVA); and, 2) that independent variables have "constant effects" across the units (or, related, the "unit homogeneity" assumption that two units will have the same value on the dependent variable when they have the same value on the explanatory variable).[3] These are very demanding assumptions, and they do not hold up when there are interaction effects among independent variables, or when there are learning or selection effects through which the outcome (or expected outcome) in one individual or group affects the behavior, treatment, or outcome of another individual or group.

For statistical methods, the possibility that there may in fact be interaction effects, selection effects, and learning can create what is known as the "ecological inference problem." Specifically, even if a statistical correlation holds up for a population, and even if the correlation is causal, it is a potential fallacy to infer that any one case in the population is causally explained by the

[1] King, Keohane, and Verba (1994). The present discussion for the most part sets aside the issue of whether this average "effect" is treated as a descriptive finding or a potentially causal relationship.

[2] In experiments, there is a stronger presumption that any difference in average outcomes between treated and untreated units is causal, and experimenters can also analyze differences in the standard deviation of outcomes between treated and untreated groups.

[3] King, Keohane, and Verba (1994: 91).

correlation that is observed at the population level. When interaction effects exist, a variable that raises the average outcome for a population may have a greater or smaller effect, or zero effect or even a negative effect, on the outcome for an individual case.

For example, in the 1960s, on the basis of statistical and other evidence, it became generally (and rightly) accepted as true that smoking increases the general prevalence of lung cancer for large groups of people. This generalization is an adequate basis for the policy recommendation that governments should discourage smoking. Yet the generalization that smoking on average increases the incidence of lung cancer does not tell us whether any one individual contracted lung cancer due to smoking.[4] Some people who smoke develop lung cancer but others do not, and some people who do not smoke develop lung cancer.[5] Scientists using statistical methods to assess epidemiological and experimental data have more recently begun to understand some of the genetic, environmental, and behavioral factors (in addition to the decision on whether to smoke) that affect the probability that a specific individual will develop lung cancer. This supports more targeted policy recommendations on whether an individual with particular genes is at especially high risk if they choose to smoke. For example, recent studies indicate that individuals with a mutation in a region on chromosome 15 will have a greatly increased risk of contracting lung cancer if they smoke (Pray 2008: 17). Even in this subgroup, however, it cannot be said with certainty that any one individual developed lung cancer because of smoking, as not every individual with this mutation gets lung cancer even if they smoke.[6]

Statistical researchers are well aware that strong assumptions are required to extend inferences from populations to individual cases, and they are

[4] This is related to the "fundamental problem of causal inference" (Holland 1986), which is that we cannot run a perfect experiment in which we rerun history, changing only one intervention or treatment, to compare an individual case that is treated in one world and untreated in the other. Here, we cannot compare the world in which an individual smokes cigarettes to the counterfactual world in which the same individual did not smoke.

[5] With other kinds of treatments or interventions, the effects on individual cases are even more uncertain. Smoking probably does not *decrease* the likelihood of lung cancer for anyone, but some medicines can cause life-saving effects in some cases, fatal allergies in others, and little or no effects in still other individuals. The language of "average effects" can be very misleading if it does not include discussion of the variance of effects: we would rightly ban a medicine that has a very small positive effect on average (say, increasing life span by a few hours) but terrible effects in a small number of cases (such as death by allergic reaction).

[6] The mechanism linking mutations on chromosome 15 to lung cancer is still under debate: the mutation could either create an indirect effect by increasing susceptibility to nicotine addiction, or a direct effect by creating molecular paths to cancer, or both (Pray 2008).

typically careful to make clear that their models do not necessarily explain individual cases (although the results of statistical studies are often oversimplified in media reports and applied to individual cases). Case study researchers in the social sciences tend to be particularly skeptical about strong assumptions regarding constant effects, unit homogeneity, and independence of cases. These researchers often think that high-order interaction effects, interdependencies among cases across space or time, and other forms of complexity are common in social life. Consequently, qualitative researchers in the social sciences typically doubt whether there are many nontrivial single-variable generalizations that apply in consistent ways across large populations of cases in society.

Case study researchers thus face the obverse of the ecological inference problem: often it is neither possible nor desirable to "generalize" from one or a few case studies to a population in the sense of developing estimates of average causal effects. Yet, at the same time, case study researchers do aspire to derive conclusions from case studies that are useful beyond the specific cases studied. Instead of seeking estimates of average effects for a population, case study researchers attempt to identify narrower "contingent generalizations" that apply to subsets of a population that share combinations of independent variables. Case study researchers thus develop "typological" or "middle range" theories about how similar combinations of variables lead to similar outcomes through similar processes or pathways. These researchers often focus on hypothesized causal mechanisms and their scope conditions, posing research questions in the following form: "Under what conditions does this mechanism have a positive effect on the outcome, under what conditions does it have zero effect, and under what conditions does it have a negative effect?"

Contingent generalizations are similar in form to the generalizations sought by statistical researchers: they apply to defined populations, they may have anomalous cases whose outcomes do not fit the generalization, and they are potentially fallible as even cases that have the expected outcome may have arrived at that outcome through mechanisms different from those associated with the theory behind the generalization. The difference is that case studies arrive at generalizations through methods that are for the most part associated with Bayesian rather than frequentist logic (see Chapter 7). Bayesian logic treats probabilities as degrees of belief in alternative explanations, and it updates initial degrees of belief (called "priors") by using assessments of the probative value of new evidence vis-à-vis alternative explanations (the updated degree of belief is known as the "posterior").

With ample cases and strong or numerous independent pieces of evidence, Bayesian and frequentist methods converge on similar conclusions, but unlike frequentism, Bayesian analysis does not need a minimum number of cases to get off the ground. Bayesianism is thus better suited to contexts in which cases are few or diverse, as is often true in the study of complex phenomena such as development.[7]

These different logics translate into differences in practice on what constitutes an acceptable generalization. Case study researchers are often happy with a generalization that holds up well for, say, five or six cases that share similar values on a half-dozen independent variables, and they are also usually curious about or troubled by individual cases that do not fit such a generalization. This is because case study researchers base their arguments on the probative value of evidence within a Bayesian framework. Within this framework, a single piece of powerful evidence can sharply discriminate between one explanation and many alternative explanations, while many pieces of weak evidence cannot support any updating unless all or most of them point in the same direction. In a frequentist framework, which treats probabilities as constituting the likelihood that a sample drawn from a population is or is not representative of the population, nothing can be said about five or six cases with seven or eight independent variables because of the "degrees of freedom" problem. Frequentists also often have little curiosity about individual cases that do not fit a correlation established through a large sample, as they expect that such outliers will occasionally happen, whether by quantum randomness or by the fact that numerous weak variables left out of a model can sometimes line up in ways that create outliers.

The different logics also lead to different ways of establishing generalizations. The above-described frequentist approach starts and ends with populations: the population is studied at the population level through the study of the full population (or the random selection of cases from the population) to make population-level claims on average effects. Case studies, in contrast, begin from within-case analysis of individual cases, or process tracing, of cases not selected at random. Process tracing uses Bayesian logic to make inferences from the evidence within a single case about alternative explanations of the outcome of that case (see Chapter 7). Depending on the results

[7] This is particularly true for rare events. Researchers sometimes closely study rare but high-consequence events such as nuclear accidents and airplane crashes, and "close calls" of events that have never happened, such as accidental use of nuclear weapons, to derive lessons for preventing rare but costly outcomes; see March, Sproull, and Tamuz (1991).

of the within-case analysis and the principle used in selecting the cases studied, case study researchers decide whether to generalize contingently (to populations that share several specified features), widely (to populations that share fewer features), or not at all. The decision on whether and how to generalize depends on the understanding that emerges from the case study regarding the mechanisms that generated the outcome of the case, and also on new and prior knowledge about the nature and prevalence of the contexts that enable those mechanisms to operate. Put another way, the study of an individual case can lead to a new understanding of causal mechanisms and the scope conditions in which they do and do not operate, and the researcher may have prior knowledge on the frequency with which the necessary scope conditions exist (and hence of the population to which the case findings are relevant).

This overall description of generalizing from case studies includes four approaches to developing generalizations: generalization from "typical" cases, generalization from most- or least-likely cases, mechanism-based generalization, and typological theorizing.[8] The sections that follow address each in turn.

## 4.3 Generalization from a "Typical" Case

A first approach to generalization from cases is to select a case that is thought to be "typical" or representative of a population (Gerring and Seawright 2008: 299–301). In the medical literature, for example, case studies are often presented as being typical of a particular disease or condition. If indeed a case is representative of a population – a key assumption – then process tracing on the case can identify or verify relationships that generalize to the population. If an existing theory predicts a population-level correlation, and statistical analysis of the relevant population exhibits the expected correlation, close study of a typical case can strengthen the inference that the correlation is causal if process tracing on the case shows the hypothesized mechanisms were indeed in operation. A typical case can also undermine causal claims if it shows that no plausible mechanisms connect the hypothesized independent variable to the outcome, or if it demonstrates that the

[8] The present discussion sets aside fuzzy-set Qualitative Comparative Analysis (fsQCA), an approach that uses fuzzy-set measures of variables, case comparisons, and Boolean algebra to find patterns on which combinations of variables relate to different outcomes; see Ragin (2006).

https://doi.org/10.1017/9781108688253.005 Published online by Cambridge University Press

mechanisms that generated the outcome were different from those initially theorized (Gerring and Seawright 2008: 299).

These inferences all depend on whether the case studied is in fact representative of the population. One way to choose a case that may be typical is to construct a statistical model and then identify a case with a small error term vis-à-vis the model, or to choose randomly from among several cases with small error terms (Gerring and Seawright 2008: 299). Added criteria for typicality could include choosing a case that is near the mean or median values on most or all variables. One problem with these criteria is that if the statistical model is mis-specified – for example, if it omits relevant variables – a case may appear to be representative when it is in fact atypical (Gerring and Seawright 2008: 300). For example, the case may include two omitted variables that occur only rarely, one of which pushes the case toward the outcome of interest and one of which inhibits or lessens the outcome, so these variables may have cancelled out each other's effects and resulted in a low error term. The case would have therefore had a low error term for reasons that would not apply to the majority of cases in the population that do not have the rare variables. One way to reduce the likelihood of this problem is to do process tracing on several cases thought to be typical.

When the population of cases is small and the hypothesized relationship involves interaction effects or different paths to the outcome that have little in common, it may be difficult or impossible to specify or identify a case that is "typical." When these conditions hold, as they often do in the study of social phenomenon, the more theory-based forms of generalization discussed herein may prove more useful than attempts to generalize from a "typical" case.

## 4.4 Generalization from Most- or Least-Likely Case Studies

The most-likely and least-likely cases approach uses extant theories and preliminary knowledge about the values of the variables in particular cases to estimate case-specific priors on how likely it is that alternative theories will prove to be good explanations of a case. A case is most-likely for a theory, or an easy test case, if we expect the theory to be a strong explanation for the case's outcome. The case is least-likely for a theory, or a tough test case, if we have reason to believe the theory should not account very well for the outcome of the case. The degree to which we can generalize from a case then depends on whether the theory passes or fails tough or easy test cases. A theory that succeeds in a least-likely case might be given broader scope conditions. For

https://doi.org/10.1017/9781108688253.005 Published online by Cambridge University Press

example, if a study shows that anarchist groups are hierarchically organized even though we should have expected them to be the least-likely kind of social organization to be hierarchical, we might conclude that hierarchy is a common feature in a wide range of social groups. Conversely, a theory that fails in a most-likely case should be assigned narrower scope conditions.[9] A theory's successful explanation of most-likely cases, or its failure to explain least-likely cases, has little impact on our estimates of its scope conditions.

Determining whether a case is most- or least-likely for a theory depends on whether the variables in the theory point strongly to an outcome, whether the variables in alternative theories point strongly to an outcome, and whether the main theory of interest and the collective alternative explanations point to the same outcome or to different outcomes. The strongest possible basis for generalizing from a case is when a theory modestly pushes toward one outcome, countervailing alternative explanations point strongly to the opposite outcome, and the first theory proves correct regarding the outcome. The strongest basis for narrowing the scope conditions of a theory exists when the theory and all the alternative explanations point strongly to the same outcome, and yet they are all wrong. Other combinations lead to different degrees of updating of scope conditions (Rapport, 2015).

An analysis of the international response to the 2014 West Africa Outbreak illustrates these issues.[10] In this outbreak the US government mobilized considerable resources – albeit later than it should have – and the UK government stepped in to assist in Sierra Leone, while France was slower to play a role and the UN system lagged.[11] There are several possible alternative explanations for the variation in these responses.[12] One possible

[9] Similarly, claims of necessity or sufficiency can be cast into doubt, in Bayesian fashion, by one or a few contrary cases.

[10] Thanks to Jennifer Widener and Michael Woolcock for providing an analysis of this example and the case codings for the USA, the WHO, and the UK (on the capacity and cohesion variables) in Table 4.1; the remaining codings and question marks are the author's. For more on this subject, see the multi-author case study series published in 2016–2017 by Princeton University's Innovations for Successful Societies research program, available at https://successfulsocieties.princeton.edu/publications/all-hands-deck-us-response-west-africa%E2%80%99s-ebola-crisis-2014-2015.

[11] Notably, each country focused on the Ebola outbreak in the African country with which it had the strongest historical ties: the USA on Liberia, the UK on Sierra Leone, and France on Guinea.

[12] We can substantially discount a fifth possible explanation – differing awareness of the problem – as it is not consistent with the variation in outcomes. The USA and the UN both had public health officials on the ground shortly after the initial cases appeared through April 2014. Although both were mistaken in thinking the epidemic had ended in April, each had kept the situation on the radar and all four governments were aware when new infections began to appear, thanks to Medicins sans frontieres (MSF).

**Table 4.1** Mobilization during 2014 Ebola outbreak: World Health Organization, United States, United Kingdom, and France

| Country or International Organization | Finance | Capacity | Authority | Cohesion | Expected Outcome (E) and Observed Outcome (O) |
|---|---|---|---|---|---|
| WHO | N | N | N | N | E: Little Mobilization<br>O: Little Mobilization |
| United States | Y | Y | Y | Y | E: Mobilization<br>O: Mobilization |
| United Kingdom | Mixed | Y | Mixed | Y | E: Slow Mobilization<br>O: Slow Mobilization |
| France | N | ? | ? | ? | E: Little Mobilization<br>O: Little mobilization |

explanation for the pattern of assistance that emerged is "Finance": the ability to summon substantial financial resources quickly. A second is "Capacity": ability to mobilize organizational resources, transportation, and medical materials rapidly. A third is "Authority": Whether there is an interagency process that allows institutions responsible for medical emergencies to work with institutions responsible for disaster response, without having to create a whole new organization for that purpose. A fourth is "Cohesion": Whether the decision to act lies within the power of one person or a few people, or whether there are many veto points.

With respect to "Finance," the USA had disaster response discretionary funds it could use to put people on the ground quickly, while the UK and France could mobilize money less easily and the UN system would have to pass the hat for contributions from member states. With respect to "Capacity," the WHO's emergency response capacity had eroded, while the USA had an Office of Foreign Disaster Assistance with a rapid response capability in place. With respect to "Authority," there was no quick way within the UN system to merge a public health or medical response (a World Health Organization matter) with a disaster response (based at the UN Office for the Coordination of Humanitarian Affairs). Finally, with respect to "Cohesion," in the USA a single decision-maker, the president, could authorize action, while the UN agencies required the assent of member-state representatives.

https://doi.org/10.1017/9781108688253.005 Published online by Cambridge University Press

In this instance, there are no strong or generalizable surprises from the most- and least-likely cases: the USA was the most-likely case for early and strong mobilization, the WHO was the least likely, and both had the expected outcomes. Had the USA failed to mobilize, or the WHO succeeded in doing so, these cases might challenge the four-factor theory of mobilization and its scope conditions.

The most interesting and strongest generalization to emerge from the international response to the 2014 Ebola outbreak is that the main bottleneck internationally was not finances or capacity, which would require financial investments to fix, but authority and cohesion, which require political attention to fix. The UK, France, and especially the USA had unused capacity in their militaries and national health systems for addressing Ebola, and the USA in particular mobilized substantial resources. However, many of these resources translated into operations only after the number of new infections per week had started to diminish. The USA deployed 3,000 troops to build 11 Ebola treatment centers in Africa, but only 28 Ebola patients received treatment at these centers, and 9 of the 11 centers never treated a single Ebola patient (Onishi 2015). In the UK, Public Health England (PHE) and the Department for International Development (DFID) coordinated in responding to Ebola, but only after initial delays that a parliamentary report attributed to over-reliance on WHO medical warning systems and DFID's inflexibility in dispersing small amounts of money early in the outbreak (House of Commons, 2016: 3). In addition, some UK health care personnel willing to volunteer for the fight against Ebola in Africa had to first negotiate leaves of absence from their respective organizations (Reece et al. 2017). A stronger and more coordinated early response would have been less costly and more effective than the slow and piecemeal responses that emerged.

## 4.5 Mechanism-Based Generalization from Cases

Typical, most-likely, and least-likely cases can provide a basis for a general claim that scope conditions should be broadened or narrowed, but they do not provide much detail on exactly how, or to what subpopulations, they might be extended, or from what subpopulations they might be withdrawn. The third, mechanism-based approach to generalizing from case studies provides some clues to this process, often by building on new theories about causal mechanisms derived from the study of individual cases.

https://doi.org/10.1017/9781108688253.005 Published online by Cambridge University Press

To understand the logic of this kind of generalization, consider two polar opposite examples, the first of which leads to very limited generalizability and the second of which leads to sweeping generalizations. In the first example, imagine that a researcher studying voter behavior finds evidence that a voter, according to the variables identified by every standard theory of voter choice (party affiliation, ideology, etc.) should have voted for candidate A, but in fact it is known that the voter chose candidate B. Imagine further that the researcher is able to ask the voter "Why did you vote for B?" and the voter replies "B is my sister-in-law." This new variable, which we might call "immediate kinship relations," provides a convincing explanation, but the mechanisms involved in the explanation suggest that it will generalize only to a very small number of cases in any election involving a large electorate.[13]

Now consider an opposite example: Charles Darwin undertook an observational study of several bird species and came up with the theory of evolutionary selection. In view of the mechanisms that this theory posits, the theory should apply to an extremely large group: all living things. Here again, the hypothesized mechanisms involved in the theory – genetic mutation, procreation, and environmental selection – provide clues on the expected scope conditions of the theory. In part, these expectations are built, in Bayesian fashion, on prior knowledge of the base rates of the enabling conditions of the theory: immediate relatives of a candidate are rare among big populations of voters, whereas living things are common.

The lessons experts drew from the early mishandling of the 2014 Ebola outbreak[14] provide a real example of generalization from an improved understanding of causal mechanisms. Here, findings on the relevant causal mechanisms are not only those concerning the medical details of the Ebola virus itself, but the interaction of the virus with local health systems, international organizations, social media, and local customs. An early opportunity to suppress the 2014 outbreak was missed because international experts did not realize that reported numbers of cases had dropped not because the outbreak had been contained, but because fearful communities had chased away health workers and sick patients were avoiding health clinics, which

[13] One could change the variable "immediate kinship relations" to a more general category of social relations (neighbors, coworkers, ethnic groups, etc.) that might apply to more cases. In cases where "ethnic voting" is common, for example, last names that are viewed as signals of ethnicity can affect voting behavior.

[14] The first cases in Guinea appeared at the very end of 2013, but for present purposes almost all the salient events unfolded in 2014 and into 2015.

they associated with high rates of death (Sack, Fink, Belluck, and Nossiter, 2014). In addition, the virus spread in part because of cultural commitments to hands-on washing of the dead, which points to the need for "culturally appropriate outreach and education" to prevent the spread of future outbreaks (Frieden et al., 2014). These findings, and not just differences in the availability of health care and quarantine technologies, help explain why Ebola spread rapidly in West Africa but not in Europe or the United States despite the arrival of infected patients in the latter regions.

Generalizations based on improved theories about causal mechanisms have two very important properties. First, they can be highly relevant for making policy decisions. For many policy decisions, we are less interested in questions such as "what is the average causal effect of X on Y in a population" than in questions such as "what will be the effect of increasing X in this particular case." Improved knowledge of how causal mechanisms work, and of the contexts in which they have positive and negative effects on the outcome of interest, is directly relevant to estimating case-specific effects.

Second, an improved understanding of causal mechanisms can allow generalizing to individual cases, and kinds of cases or contexts, that are different from or outside of the sample of the cases studied. This is a very important property of theoretical understandings derived from the close observation of causal mechanisms in individual cases, as both statistical studies and artificial intelligence algorithms are often weak at "out of sample" predictions. A powerful example here is the development of an effective "cocktail" of drugs to treat HIV-AIDS. This medical advance was greatly fostered by the close study of individual patients who responded far better to treatments than other patients. Researchers concluded upon close examination of such patients that administration of a combination of drugs earlier in the progression of the disease than previous experimental treatments could keep it in check (Schoofs, 1998). This illustrates that when a researcher comes up with a new theory or explanation from the study of a case, their new understanding of the hypothesized mechanisms through which the theory operates can itself give insights into the expected scope conditions of the theory, as in the above-mentioned "sister-in-law" and Darwin examples.

While researchers might derive new understandings of causal mechanisms from many types of case studies, two kinds of case selection are particularly oriented toward developing new understandings of mechanisms and their scope conditions: studies of "deviant" (or outlier) cases, and studies of cases

that have high values on an independent variable of interest.[15] Deviant cases, or cases with an unexpected outcome or a high error term relative to extant theories, are good candidates for the purpose of looking inductively for new explanations or omitted variables. In these cases, new insights and theories may arise from the inductive use of process tracing to connect "clues" – pieces of evidence that do not on first examination fit into extant theories – in a new explanation.[16]

An interesting and important dilemma here concerns decisions on whether to stop trial experiments on medical or other treatments sooner than planned when the early subjects undergoing the treatment show signs of either catastrophic failures or remarkable successes. Continuing a trial after a treatment has shown signs of being powerfully effective can be unethical as it delays treatment of other individuals or communities who might benefit. Even worse, continuing a trial treatment after catastrophic outcomes arise in early cases can cost lives. Much of the discussion of this issue in the medical literature warns against premature termination of medical experiments, regardless of unexpectedly good or bad early results, due to the frequentist argument that small samples can be unrepresentative and do not allow powerful conclusions. There is indeed a risk that trials stopped early for benefit might catch the observed treatment effect at a "random high," which later can yield to a "regression to the truth effect" in subsequent trials or clinical use (Montori et al., 2005). Yet qualitative evidence from individual cases can provide additional analytical leverage over decisions on whether to continue experiments after strong early results, particularly when that evidence, combined with existing expert knowledge, strongly illuminates the causal mechanisms at work. Experts on clinical trials have thus noted that "formal statistical methods should be used as tools to guide decision-making rather than as hard rules" (Sydes et al., 2004: 60) and that "predefined statistical stopping boundaries for benefit provide a useful objective guideline, but the reality of making wise judgements on when to stop involves an evaluation of the totality of evidence available" (Pocock, 2006: 516).

Bayesian logic and process tracing provide a useful perspective on this issue. As noted, whereas frequentism treats probabilities as representing the likelihood that a sample is representative of a population, Bayesians view probability as representing degrees of belief in different explanations.

[15] Recent research on case selection has placed renewed emphasis on the value of deviant and "high on the independent variable" cases as sources of insights on causal mechanisms (Seawright, 2016).

[16] A large error term can also arise from the combined effects of many different weak variables, or from measurement error, rather than from one or a few strong omitted variables.

Consequently, when evidence is uniquely consistent with one explanation, Bayesians can update their confidence in alternative explanations even with small numbers of cases. In medical applications, this involves looking at process-tracing evidence on *why* a treatment succeeded or failed, not just whether it succeeded or failed. While much of the thinking behind clinical trials still reflects a frequentist outlook, a more Bayesian and process-tracing approach has been influential in epidemiology and experimental medicine as well. Early on in the debates on the relationship between smoking and cancer, the English epidemiologist Sir Austin Bradford Hill developed nine criteria for assessing evidence on a potential causal relationship between a presumed cause and an observed effect. These include process-tracing types of criteria, such as the specificity of the observed relationship, the temporal precedence of the cause over the effect, and the existence of a plausible theorized mechanism linking the cause and the effect. As a later study of Hill's criteria concluded: "Whereas a trial is often open to the objection that it is an anomaly or not generalizable, if we supplement the evidence from the trial with strong mechanistic and parallel evidence, it becomes increasingly difficult to question the results of the study and its applicability to a wider target population" (Howick, Glasziou, and Aronson, 2009: 193).

An example here concerns the early application of chimeric antigen receptor T-cell (CAR-T) therapy. In CAR-T therapy, physicians alter a patient's T-cells (a type of white blood cell critical to the immune system) so that these T-cells can better target and destroy cancer cells. The physicians then introduce the altered T-cells back into the patient's body. Of the first patients with ordinarily fatal cancers given this experimental treatment, three had complete remissions, four improved without a full remission, one improved and then relapsed, and two showed no effect. While these early results included too few cases for any strong conclusion using frequentist statistics, they looked promising given the extremely low remission rates of untreated patients with the kinds of cancers included in the initial study, and research on CAR-T therapy continued.

The most revealing case arose when doctors chose to administer CAR-T therapy in 2012 to Emily Whitehead, a young patient with a likely terminal case of Acute Lymphoblastic Leukemia. Like some previous CAR-T patients, within a few days Emily developed life-threatening immune response symptoms, including a fever of 105 degrees, and appeared to be hours away from death. Fortunately, her doctors quickly found that the cause was an elevation of cytokines, inflammatory factors secreted by T-cells and their target cells. Emily had one cytokine in particular, IL-6, that was 1,000 times higher than

normal. In a sense, given her doctors' already well-developed understanding of their therapeutic approach, this showed that the CAR-T process was working: the chimeric T-cells were targeting and destroying cancer cells at an astonishing rate. Yet the associated side effect of inflammation might have killed Emily, as it had a previous patient named Jesse Gelsinger. Luckily, one of Emily's doctors knew of a recently approved drug that blocks IL-6, and Emily experienced a remarkably quick and full recovery once she received this drug. Seven years later, she remained cancer-free (Mukherjee, 2019).

This example demonstrates that the efficacy and generalizability of an intervention should rely not only on the number of successes or failures and frequentist statistical assumptions about sampling, but also on Bayesian inference, prior theoretical knowledge, and process-tracing evidence. Here, despite the small number of prior cases, the results were striking: deadly in some cases, remarkably curative for some who survived the inflammatory response. Emily's case provided the key process-tracing clue regarding the "cytokine storm" that was threatening patients. Fortunately, a drug was at hand to treat her particular IL-6 cytokine spike, and doctors used their prior causal knowledge to decide to administer this drug. Emily's recovery spurred further CAR-T research, and while not every patient has benefited in trials and several challenges remain, the therapy continues to show promise. Yet given the frequentist tilt of extant practices in medical research, the future of CAR-T therapy hinged on Emily's personal outcome to a far larger degree than it should have. As one physician later commented (Rosenbaum, 2017: 1314):

> anecdote can easily break a field rather than make it: the death of Jesse Gelsinger in a trial at Penn had set the field of gene therapy back at least a decade. And as both June and Stephan Grupp, the Children's Hospital oncologist and principal investigator of the CART-19 trial in children, emphasized, had Emily died, the CAR-T field would probably have died with her.

In addition to studying cases with remarkable outcomes on the dependent variable, the study of cases with high values on an independent variable of interest can contribute to better and generalizable understandings of causal mechanisms. This is often the intuition behind selecting cases that have high value on both an independent variable and the dependent variable. An example here is a study of "hybrid" microfinance organizations, or commercial organizations that combine elements of profit-making lending and development-oriented lending, by Julie Battilana and Silvia Dorado. These authors chose two such organizations in Bolivia that they knew to be

https://doi.org/10.1017/9781108688253.005 Published online by Cambridge University Press

"pioneering" and high-performing in order to carry out a "comparative inductive study" (Battilana and Dorado, 2010:1435) of the factors behind their success. They concluded from close study of these two organizations that their innovative hiring and socialization processes accounted for their high portfolio growth. The authors suggest that this finding is relevant to hybrid organizations more generally, although they also note "limits to the influence of hiring and socialization policies in mitigating tensions between institutional logics within organizations" (Battilana and Dorado, 2010: 1420).

Of course, researchers can make mistakes in either over-generalizing or under-generalizing the expected scope conditions that emerge from their understanding of a new theory. For this reason, while researchers may have warrant for making claims on the scope conditions of new theories derived from cases, these claims must remain provisional pending testing in other cases. Researchers should be particularly careful of selecting "best practices" cases on the basis of performance or outcomes, or selecting on the dependent variable, and then making inferences on the practices in these cases as the causes of high performance. If a population is large, some units may perform well even over long periods of time just by chance. Researchers have often claimed to have found the best practices that underlie unusually good performance in companies' stock market strategies and management practices, for example, only to find later that the same companies later experienced average or even below average performance, exhibiting regression toward the mean.[17]

## 4.6 Typological Theorizing and Generalization

The fourth approach to generalization from case studies, typological theorizing, systematically combines process tracing and small-N comparisons. The goal is to develop a theory on different combinations of independent variables, or types, so that contingent generalizations can be made about the processes and outcomes of cases within each type.[18] To develop and test these

[17] A well-known example here is the popular management book *In Search of Excellence* (Peters and Waterman, 1982), which studied high-performing businesses and claimed to have found the common principles that led to their above-average returns. Within a few years of the book's publication, most of the businesses that were the basis of the study experienced average or poor returns.

[18] Qualitative Comparative Analysis (QCA) is similar to typological theorizing in that it focuses on cases as combinations of variables, but as traditionally practiced QCA relies on cross-case comparisons rather than within-case analysis. More recently, some QCA methodologists have advocated using QCA for the purpose of selecting cases for process tracing, which is more similar to typological theorizing (Schneider and Rohlfing 2013).

contingent generalizations, researchers first build a typological theory, starting deductively and then iterating between their initial theoretical understanding of the phenomenon they are studying and their initial knowledge of the measures of the variables in the cases in the relevant population. Once they have built a typological theory using this initial knowledge, the researchers can use it to choose which cases they will study, and then they can use process tracing (see Chapter 7) to study those cases.

While a full discussion of typological theorizing is beyond the scope of this chapter,[19] the paragraphs that follow outline a process for developing typological theories. As an illustrative example, the discussion considers the puzzle of why, in response to epidemics such as Ebola or flu, governments sometimes resort to isolation strategies while at other times they employ quarantines. Isolation involves treating and limiting the movement of symptomatic patients suspected of having a contagious disease, while quarantines seek to limit the movement into and out of designated areas (including neighborhoods or whole cities) of individuals who may have been exposed to an illness but are not themselves symptomatic. Isolation is uncontroversial, while quarantines raise more difficult issues regarding civil liberties. Quarantines can also create unintended consequences by inhibiting patients who might be sick from seeking care, or motivating individuals to flee from high-infection quarantined areas to low-infection areas, possibly spreading the epidemic in the process. For present purposes of illustrating a typological theory, however, I focus not on the policy question of when quarantines might be efficacious, or the ethical question of when they might be justified, but the political question of when they are attempted.

To build a typological theory, the researcher first defines or conceptualizes the outcome of interest (the dependent variable) and decides how to measure this outcome. Often in typological theories the dependent variable is categorized by nominal measures (such as "democracy" and "non-democracy"), but it can also be categorized by ordinal measures (such as high, medium, and low levels of growth in the percentage of children attending school), or by conceptual typologies (such as combinations of variables that constitute three types of "welfare capitalism" (Esping-Andersen, 1990)). In our example of isolation

One advantage of QCA is that it allows the derivation of two different measures relating to generalization. The first measure, "consistency," assesses the degree to which cases that share a condition or a combination of variables have the same outcome. The second measure, "coverage," estimates the degree to which any variable or combination covers the total of instances of the outcome of interest. This is a measure of the importance of the variable or combination (Ragin, 2006).

[19] See George and Bennett (2005) and Bennett (2013).

https://doi.org/10.1017/9781108688253.005 Published online by Cambridge University Press

versus quarantine, there are gradations of both (How many symptoms qualify a patient for isolation? How geographically broad or narrow is a quarantine and does it allow many or few exceptions for work or family reasons?), but the overall conceptual difference between isolation and quarantine is clear. For present purposes, the discussion therefore uses a simple dichotomized dependent variable of isolation versus quarantine, but subsequent research could consider gradations and kinds of isolations and quarantines.

Second, the researcher draws on existing theories to identify the key independent variables from individual theories, or constituent theories that relate to the outcome of interest. By convention, these independent variables constitute the columns in a table laying out the typological theory, while the individual cases (or clusters of cases with the same combination of independent variables, or "types") constitute the rows in the typological table. In our example I offer three independent variables that may affect choices between isolation strategies and quarantines. First, airborne epidemics, which typically spread quickly, are more likely to be subject to quarantine than those transmitted only by direct bodily contact. This may even be a nearly sufficient condition for quarantines. Second, isolation is more likely when a country has a high-capacity health care system that can treat a large number of individuals. Third, quarantines are a more tempting option when individuals in the quarantined area have few transportation or other options for escaping the quarantine area. Additional variables may matter as well, such as levels of social media, levels of trust or distrust in the government and the health system, and state capacity for coercion, but for illustrative purposes the present example includes only three independent variables and treats each as dichotomous.

Third, the researcher builds a table – a "typological space" (sometimes called a "possibility space" or a "property space" in the philosophy of logic) of all the possible combinations of the independent variables of the constituent theories.[20] Because a typological space becomes combinatorially more complex with additional variables and finer levels of measurement of these variables, for the purpose of presenting and thinking through the typological table, researchers typically include six or fewer independent variables and use nominal, dichotomous, or trichotomous measures of these variables. Researchers can relax the simplifications on the number and measurement

[20] For discussion and a compilation of examples of theoretical typologies, see Collier, Laporte, and Seawright (2012).

https://doi.org/10.1017/9781108688253.005 Published online by Cambridge University Press

**Table 4.2** A typological theory on government choices of isolation versus quarantine strategies in epidemics

| Case | Air or Direct Transmission | High or Low Health Care Capacity | High or Low Ability to Escape Quarantine | Outcome: Expected (E) and Observed (O) |
|---|---|---|---|---|
| SARS 2003 in Taiwan, Canada, | Air | H | H | Unclear Prediction; Quarantine (O) |
| SARS 2003 in Hong Kong, Singapore | Air | H | L | Quarantine (E) Quarantine (O) |
| SARS 2003 in Vietnam | Air | L | H | Quarantine (E) Quarantine (O) |
| SARS 2003 in China | Air | L | L | Quarantine (E) Quarantine (O) |
| Ebola 2013–2015 in the United States, EU countries | Direct | H | H | Isolation (E) Isolation (O) |
| No cases | Direct | H | L | Isolation (E) |
| Ebola 2013–2015 in Guinea, Liberia, Sierra Leone | Direct | L | H | Unclear Prediction; Liberia attempted quarantine, others did not |
| No cases | Direct | L | L | Unclear Prediction |

of variables as they move from the simplified typological theory to the within-case analysis of individual cases. In our example, with three dichotomous variables, we have two to the power of three or eight possible combinations. These are outlined in Table 4.2.

Fourth, the researcher deductively thinks through how each combination of variables might interact and what the expected outcome should be for each row. This is the step at which the researcher integrates the constituent theories that created the typological space into a single typological theory that provides the expected outcome for every combination of variables. In practice, a typological theory is rarely fully specified, as the researcher may

https://doi.org/10.1017/9781108688253.005 Published online by Cambridge University Press

lack a strong theoretical prior for every possible combination of the independent variables. Still, it is useful to think through possible interactions and specify expected outcomes deductively to the extent possible. Table 4.2 identifies the expected outcome for combinations that lead to clear and strong predictions on outcomes, such as combinations where all three independent variables point to the same expected outcome and interaction effects are unlikely. Table 4.2 codes a question mark for combinations in which the independent variables push toward different outcomes.

Fifth, after this deductive construction of the first draft of the typological theory, the researcher can use their preliminary empirical knowledge of extant historical cases to classify these cases into their respective types or rows. This stage allows for some iteration between the researcher's preliminary theoretical expectations and their initial knowledge of the empirical cases. Quick initial comparisons of the cases might lead to revisions to the theoretical typology and/or to the remeasurement and reclassification of cases. For example, if cases are in the same row – that is, they have fully similar combinations of the values of the independent variables – but they have different outcomes, they pose anomalies for the emerging theory. A quick examination of these cases might lead to revisions in the typology or the measurement of the variables in the cases in question, or deeper process tracing may be necessary to analyze why the cases have different outcomes. The example in Table 4.2 includes countries that had a significant number of SARS cases in 2003 or Ebola cases in 2013–2015, and it also includes some countries that had a few Ebola cases but public debates over a possible quarantine. The codings are based on very limited and preliminary knowledge of the values of the variables in each case, particularly the measurement of the ability of individuals to escape quarantined areas.

After iterating between the typological theory and the classification of extant cases to resolve all the discrepancies that can be addressed quickly and easily with the benefit of secondary sources, the researcher can undertake the sixth step: using the refined typological theory to select cases for deeper research that uses process tracing. The refined typological theory makes it easy to assess which cases fit various comparative research designs and inferential purposes: most-similar cases (cases that differ on one independent variable and on the outcome), least-similar cases (cases with the same outcome and only one independent variable in common), deviant cases (cases without the predicted outcome), cases with a high value on one independent variable, and typologically similar cases (cases in the same type or row and with the same outcome). In this example, interesting cases

worth studying are those of Liberia, Sierra Leone, and Guinea. The theory does not make a strong prediction for the combination of variables evident in the cases of Sierra Leone, Guinea, and Liberia in 2013–2015 because the high ability of individuals to escape quarantine and the low capacity to isolate and treat patients push in opposite directions. Comparisons among these cases could prove fruitful in understanding why only Liberia attempted a quarantine.

Vietnam is also an interesting case worthy of study, as it was fairly successful in containing SARS despite limited health resources (Rothstein et al., 2003: 107). This makes it a least-likely case that succeeded. Canada and Taiwan are worthy of study as well, as the theory does not give a strong prediction on how countries with high health care capacity (and here, strong democratic cultures) would respond to airborne epidemics, and both countries resorted to quarantines.

This is a "building block" approach in several senses: it builds on theories about individual variables or mechanisms, theorizes about different combinations of these variables, uses individual case studies to validate the theorization on each combination of variables or "type" of case, and cumulatively charts out different types or paths to the outcome of interest. If there are limited interaction effects, individual variables, or even combinations of variables, will behave similarly across types, but typological theorizing does not presume or require such constant or simple interaction effects. Its strongest generalizations focus on the cases within each type. This prioritizes theoretical intension – making strong statements about well-defined subtypes that cover relatively few cases – while it sacrifices some degree of parsimony, as each combination or path can have its own explanation. Typological theorizing does not necessarily aspire to single-variable generalizations that apply to the whole population, but if such generalizations exist, it can still uncover them. In our example, both the theory and the extant cases suggest that quarantines are far more likely for airborne epidemics.

## 4.7 Generalizing – Carefully and Contingently – from Cases

Researchers in both the qualitative and quantitative traditions are rightly cautious about generalizing from individual case studies to broad populations. Case studies are not optimal for generalizing in the sense of estimating average effects for a population, as statistical studies aim to do. In addition, when process tracing reveals that the outcome in a case was due to

https://doi.org/10.1017/9781108688253.005 Published online by Cambridge University Press

mechanisms whose enabling conditions are rare or unique, little or no generalization beyond the case is possible. Even when findings do generalize from individual cases, it can be difficult to identify exactly the scope conditions in which they apply.

Yet case studies contribute to forms of generalization that are different from average population-level effects and that are pragmatically useful for policymakers. Cases that are typical, most-likely, least-likely, deviant, and high on the value of a particular independent variable can all contribute to various forms of generalization even if they do not always provide clear guidelines on the scope conditions for generalizations. And sometimes cases do allow inferences about scope conditions – the clearer understanding of causal mechanisms that often emerges from process tracing can provide information on the conditions under which these mechanisms operate, and prior knowledge can indicate how common those conditions are. Just as a case study can uncover causal mechanisms that are relatively unique, it can also identify mechanisms that prove generalizable to large populations. In addition, typological theorizing can develop contingent generalizations about cases that share combinations of variables. Researchers can also develop cumulatively better knowledge of a phenomenon as they build upon and revise typological theories through the study of additional or subsequent cases.

These forms of generalization from case studies are Bayesian in the sense that they depend on prior theoretical knowledge and knowledge about the prevalence of the scope conditions thought to enable causal mechanisms to operate. Prior knowledge on both how causal mechanisms operate and where/under what conditions they operate can be updated through the study of individual cases. As prior knowledge is usually incomplete, however, generalization from cases is potentially fallible. Researchers can make the mistake of either over-generalizing or under-generalizing from cases. Process-tracing research on additional cases, as well as statistical studies of newly modeled mechanisms, can further test whether generalizations about causal mechanisms hold, and whether they need to be modified. Careful generalizations from case studies can thus contribute to cumulating policy-relevant knowledge about causal processes and the conditions under which they operate.

## References

Battilana, J. and Dorado, S. (2010) "Building sustainable hybrid organizations: The case of commercial microfinance organizations," *Academy of Management Journal*, 53(6), 1419–1440.

Bennett, A. (2013) "Causal mechanisms and typological theories in the study of civil conflict" in Checkel, J. (ed.) *Transnational dynamics of civil war*. New York: Cambridge University Press, pp. 205–230.

Collier, D., Laporte, J., and Seawright, J. (2012) "Putting typologies to work: Concept formation, measurement, and analytic rigor," *Political Research Quarterly*, 65(1), 217–232.

Esping-Andersen, G. (1990) *The three worlds of welfare capitalism*. Hoboken, NJ: John Wiley and Sons.

Flyvbjerg, B. (2006) "Five misunderstandings about case-study research," *Qualitative Inquiry*, 12(2), 219–245.

Frieden, T., Damon, I., Bell, B., Kenyon, T., and Nichol, S. (2014) "Ebola 2014 – New challenges, new global response and responsibility," *New England Journal of Medicine*, 371(13), 177–1180.

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

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

Holland, P. W. (1986) "Statistics and causal inference," *Journal of the American Statistical Association*, 81(396), 945–960.

House of Commons International Development Committee. (2016) "Ebola: Responses to a public health emergency," Second Report of Session 2015–2016.

Howick, J., Glasziou, P., and Aronson, K. (2009) "The evolution of evidence hierarchies: What can Bradford Hill's 'guidelines for causation' contribute?" *Journal of the Royal Society of Medicine*, 102(5), 186–194.

King, G., Keohane, R., and Verba, S. (1994) *Designing social inquiry*. Princeton, NJ: Princeton University Press.

March, J., Sproull, L., and Tamuz, M. (1991) "Learning from samples of one or fewer," *Organization Science*, 2(1), 1–13.

Montori, V. M., Devereaux, P. J., Adhikari, N. K., et al. (2005) "Randomized trials stopped early for benefit: A systematic review," *JAMA*, 294(17), 2203–2209.

Mukherjee, S. (2019) "The promise and price of cellular therapies," *The New Yorker*, July 15.

Onishi, N. (2015) "Empty Ebola clinics in Liberia are seen as misstep in US relief effort," *The New York Times*, April 11.

Peters, T. and Waterman, R. (1982) *In search of excellence: Lessons from America's best-run companies*. New York: Harper and Row.

Pocock, S. J. (2006) "Current controversies in data monitoring for clinical trials," *Clinical Trials*, 3(6), 513–521.

Pray, L. (2008) "Genes, smoking, and lung cancer," *Nature Education*, 1(1), 73.

Ragin, C. (2006) "Set relations in social research: Evaluating their consistency and coverage," *Political Analysis*, 14(3), 291–310.

Ragin, C. (2008) *Redesigning social inquiry: Fuzzy sets and beyond*. Chicago, IL: University of Chicago Press.

Rapport, A. (2015) "Hard thinking about hard and easy cases in security studies," *Security Studies*, 24(3), 431–465.

Reece, S., Brown, C. S., Dunning, J., Chand, M. A., Zambon, M. C., and Jacobs. M. (2017) "The UK's multidisciplinary response to an Ebola epidemic," *Clinical Medicine*, 17(4), 332–337.

Rosenbaum, L. (2017) "Tragedy, perseverance, and chance – The story of CAR-T therapy," *The New England Journal of Medicine*, 377(14), 1313–1315.

Rothstein, M. A., Alcalde, M. G., Elster, N. R., et al. (2003) *Quarantine and isolation: Lessons learned from SARS*. A Report to the Centers for Disease Control and Prevention.

Sack, K., Fink, S., Belluck, P., and Nossiter, A. (2014) "How Ebola roared back," *The New York Times Magazine*, December 29.

Schneider, C. and Rohfling, I. (2013) "Combining QCA and process tracing in set-theoretic multimethod research," *Sociological Methods and Research*, 42(4), 559–597.

Schoofs, M. (1998) "The Berlin patient," *The New York Times Magazine*, June 21.

Seawright, J. (2016) "The case for selecting cases that are deviant or extreme on the independent variable," *Sociological Methods & Research*, 45(3), 493–525.

Sydes, M. R., Spiegelhalter, D. H., Altman, D. G., Babiker, A. B., and Parmar, M. K. B. (2004) "Systematic qualitative review of the literature on data monitoring committees for randomized controlled trials," *Clinical Trials*, 1(1), 60–79.