

PURPOSEFUL SAMPLING

Perhaps nothing better captures the difference between quantitative and qualitative methods than the different logics that undergird sampling approaches. Qualitative inquiry typically focuses in depth on relatively small samples, even single cases ($n = 1$), selected *purposefully*. Quantitative methods typically depend on larger samples selected randomly. Not only are the techniques for sampling different, but the very logic of each approach is unique because the purpose of each strategy is different.

The logic and power of probability sampling depends on selecting a truly random and statistically representative sample that will permit confident generalization from the sample to a larger population. The purpose is generalization.

The logic and power of purposeful sampling lies in selecting *information-rich cases* for study in depth. Information-rich cases are those from which one can learn a great deal about issues of central importance to the purpose of the research, thus the term *purposeful* sampling. For example, if the purpose of an evaluation is to increase the effectiveness of a program in reaching lower-socioeconomic groups, one may learn a great deal more by focusing in depth on understanding the needs, interests, and incentives of a small number of carefully selected poor families than by gathering standardized information from a large, statistically representative sample of the whole program. The purpose of purposeful sampling is to select information-rich cases whose study will illuminate the questions under study.

There are several different strategies for purposefully selecting information-rich cases. The logic of each strategy serves a particular evaluation purpose.

(1) *Extreme or deviant case sampling*. This approach focuses on cases that are rich in information because they are unusual or special in some way. Unusual or special cases may be particularly troublesome or especially enlightening, such as outstanding successes or notable failures. If, for example, the evaluation was aimed at gathering data help a national program reach more clients, one might compare a few project sites that have long waiting lists with those that have short waiting lists. If staff morale was an issue, one might study and compare high-morale programs to low-morale programs.

The logic of extreme case sampling is that lessons may be learned about unusual conditions or extreme outcomes that are relevant to improving more typical programs. Let's suppose that we are interested in studying a national program with hundreds of local sites. We know that many programs are operating reasonably well, even quite well, and that other programs verge on being disasters. We also know that most programs are doing "okay." This information comes from knowledgeable sources who have made site visits to enough programs to have a basic idea about what the variation is. The question is this: How should programs be sampled for the study? If one wanted to precisely document the natural variation among programs, a random sample would be appropriate, preferably a random sample of sufficient size to be truly representative of and permit generalizations to the total population of programs. However, some information is already available on what program variation is like. The question of more immediate interest may concern *extreme cases*. With limited resources and limited time an evaluator might learn more by intensively studying one or more examples of really poor programs and one or more examples of really excellent programs. The evaluation focus, then, becomes a question of understanding under what conditions programs get into trouble and under what conditions programs exemplify excellence. It is not even necessary to randomly sample poor programs or excellent programs. The researchers and intended users involved in the study think through *what cases they could learn the most from* and those are the cases that are selected for study.

In a single program the same strategy may apply. Instead of studying some representative sample of people in the setting, the evaluator may focus on studying and understanding selected cases of special interest, for example, unexpected dropouts or outstanding successes. In many instances more can be learned from intensively studying extreme or unusual cases than can be learned from statistical depictions of what the average case is like. In other evaluations detailed information about special cases can be used to supplement statistical data about the normal distribution of participants.

Ethnomethodologists use a form of extreme case sampling when they do their field experiments. Ethnomethodologists are interested in everyday experiences of routine living that depend on deeply understood, shared understandings among people in a setting (see Chapter 3). One way of exposing these implicit assumptions and norms on which everyday life is based is to create disturbances that

deviate from the norm. Observing the reactions to someone eating like a pig in a restaurant and then interviewing people about what they saw and how they felt would be an example of studying a deviant sample to illuminate the ordinary.

The Peters and Waterman (1982) best-selling study of "America's best run companies," *In Search of Excellence*, exemplifies the logic of purposeful, extreme group sampling. Their study was based on a sample of 62 companies "never intended to be perfectly representative of U.S. industry as a whole ... [but] a list of companies considered to be innovative and excellent by an informed group of observers of the business scene" (Peters and Waterman, 1982: 19).

Another excellent example of extreme group sampling is Angela Browne's (1987) study, *When Battered Women Kill*. She conducted in-depth studies of the most extreme cases of domestic violence to elucidate the phenomenon of battering and abuse. The extreme nature of the cases presented are what render them so powerful. Browne's book is an exemplar of qualitative inquiry using purposeful sampling for applied research.

(2) *Intensity sampling*. Intensity sampling involves the same logic as extreme case sampling but with less emphasis on the extremes. An intensity sample consists of information-rich cases that manifest the phenomenon of interest intensely (but not extremely). Extreme or deviant cases may be so unusual as to distort the manifestation of the phenomenon of interest. Using the logic of intensity sampling, one seeks excellent or rich examples of the phenomenon of interest, but not unusual cases.

Heuristic research uses intensity sampling. Heuristic research draws explicitly on the intense personal experiences of the researcher, for example, experiences with loneliness or jealousy. Coresearchers who have experienced these phenomena intensely also participate in the study (see Chapter 3). The heuristic researcher is not typically seeking pathological or extreme manifestations of loneliness, jealousy, or whatever phenomenon is of interest. Such extreme cases might not lend themselves to the reflective process of heuristic inquiry. On the other hand, if the experience of the heuristic researcher and his or her coresearchers is quite mild, there won't be much to study. Thus the researcher seeks a sample of sufficient intensity to elucidate the phenomenon of interest.

The same logic applies in a program evaluation. Extreme successes or unusual failures may be discredited as being too extreme or un-

usual for gaining information. Therefore, the evaluator may select cases that manifest sufficient intensity to illuminate the nature of success or failure, but not at the extreme.

Intensity sampling involves some prior information and considerable judgment. The researcher must do some exploratory work to determine the nature of the variation in the situation under study. One can then sample intense examples of the phenomenon of interest.

(3) Maximum Variation sampling. This strategy for purposeful sampling aims at capturing and describing the central themes or principal outcomes that cut across a great deal of participant or program variation. For small samples a great deal of heterogeneity can be a problem because individual cases are so different from each other. The maximum variation sampling strategy turns that apparent weakness into a strength by applying the following logic: Any common patterns that emerge from great variation are of particular interest and value in capturing the core experiences and central, shared aspects or impacts of a program.

How does one maximize variation in a small sample? One begins by identifying diverse characteristics or criteria for constructing the sample. Suppose a statewide program has project sites spread around the state, some in rural areas, some in urban areas, and some in suburban areas. The evaluation lacks sufficient resources to randomly select enough project sites to generalize across the state. The evaluator can at least be sure that the geographical variation among sites is represented in the study.

When selecting a small sample of great diversity, the data collection and analysis will yield two kinds of findings: (1) high-quality, detailed descriptions of each case, which are useful for documenting uniqueness, and (2) important shared patterns that cut across cases and derive their significance from having emerged out of heterogeneity.

The same strategy can be used within a single program in selecting individuals for study. By including in the sample individuals the evaluator determines have had quite different experiences, it is possible to more thoroughly describe the variation in the group and to understand variations in experiences while also investigating core elements and shared outcomes. The evaluator using a maximum variation sampling strategy would not be attempting to generalize findings to all people or all groups but would be looking for information that elucidates programmatic variation and significant common patterns within that variation.

(4) Homogeneous samples. In direct contrast to maximum variation sampling is the strategy of picking a small homogeneous sample. The purpose here is to describe some particular subgroup in depth. A program that has many different kinds of participants may need in-depth information about a particular subgroup. For example, a parent education program that involves many different kinds of parents may focus a qualitative evaluation on the experiences of single-parent female heads of household because that is a particularly difficult group to reach and hold in the program.

Focus group interviews are typically based on homogeneous groups. Focus group interviews involve conducting open-ended interviews with groups of five to eight people on specially targeted or focused issues. The use of focus groups in evaluation will be discussed at greater length in the chapter on interviewing. The point here is that sampling for focus groups typically involves bringing together people of similar backgrounds and experiences to participate in a group interview about major program issues that affect them.

(5) Typical case sampling. In describing a program or its participants to people not familiar with the program it can be helpful to provide a qualitative profile of one or more "typical" cases. These cases are selected with the cooperation of key informants, such as program staff or knowledgeable participants, who can help identify what is typical. It is also possible to select typical cases from survey data, a demographic analysis of averages, or other programmatic data that provide a normal distribution of characteristics from which to identify "average" examples. Keep in mind that the purpose of a qualitative profile of one or more typical cases is to describe and illustrate what is typical to those unfamiliar with the program—not to make generalized statements about the experiences of all participants. The sample is illustrative not definitive.

When entire programs or communities are the unit of analysis, it is also possible to sample somewhat typical cases. Again, the study of such typical programs does not, of course, permit generalizations in any rigorous sense. It does, however, mean that the processes and effects described for the typical program need not be dismissed as peculiar to "poor" sites or "excellent" sites. When the typical site sampling strategy is used, the site is specifically selected because it is not in any major way atypical, extreme, deviant, or intensely unusual. This strategy is often appropriate in sampling villages for community development studies in Third World countries. A study of a typical

village illuminates key issues that must be considered in any development project aimed at this kind of village.

Decision makers may have made their peace with the fact that there will always be some poor programs and some excellent programs, but the programs they really want more information about are those run-of-the-mill programs that are "hard to get a handle on." It is important, when using this strategy, to attempt to get broad consensus about which programs are "typical." If a number of such programs are identified, only a few can be studied, and there is no other basis for selecting among them purposefully, then it is possible to randomly select from among all "typical" programs identified to select those few typical cases that actually will be included in the study.

(6) *Stratified purposeful sampling.* It is also clearly possible to combine a typical case sampling strategy with others, essentially taking a *stratified purposeful sample* of above average, average, and below average cases. This is less than a full maximum variation sample. The purpose of a stratified purposeful sample is to capture major variations rather than to identify a common core, although the latter may also emerge in the analysis. Each of the strata would constitute a fairly homogeneous sample. This strategy differs from stratified random sampling in that the sample sizes are likely to be too small for generalization or statistical representativeness.

(7) *Critical case sampling.* Another strategy for selecting purposeful samples is to look for critical cases. Critical cases are those that can make a point quite dramatically or are, for some reason, particularly important in the scheme of things. A clue to the existence of a critical case is a statement to the effect that "if it happens there, it will happen anywhere," or, vice versa, "if it doesn't happen there, it won't happen anywhere." The focus of the data gathering in this instance is on understanding what is happening in that critical case. Another clue to the existence of a critical case is a key informant observation to the effect that "if that group is having problems, then we can be sure all the groups are having problems."

Looking for the critical case is particularly important where resources may limit the evaluation to the study of only a single site. Under such conditions it makes strategic sense to pick *the* site that would yield the most information and have the greatest impact on the development of knowledge. While studying one or a few critical cases does not technically permit broad generalizations to all possible cases,

logical generalizations can often be made from the weight of evidence produced in studying a single, critical case.

Physics provides a good example of such a critical case. In Galileo's study of gravity he wanted to find out if the weight of an object affected the rate of speed at which it would fall. Rather than randomly sampling objects of different weights in order to generalize to all objects in the world, he selected a critical case—the feather. If in a vacuum, as he demonstrated, a feather fell at the same rate as some heavier object (a coin), then he could logically generalize from this one critical case to all objects. His findings were enormously useful *and* credible.

There are many comparable critical cases in social science research—if one is creative in looking for them. For example, suppose national policymakers want to get local communities involved in making decisions about how their local program will be run, but they aren't sure that the communities will understand the complex regulations governing their involvement. The first critical case is to evaluate the regulations in a community of well-educated citizens; if they can't understand the regulations, then less-educated folks are sure to find the regulations incomprehensible. Or conversely, one might consider the critical case to be a community consisting of people with quite low levels of education: "If they can understand the regulations, anyone can."

Identification of critical cases depends on recognition of the key dimensions that make for a critical case. A critical case might be indicated by the financial state of a program; a program with particularly high or particularly low cost-per-client ratios might suggest a critical case. A critical case might come from a particularly difficult program location. If the funders of a new program are worried about recruiting clients or participants into a program, it may make sense to study the site where resistance to the program is expected to be greatest to provide the most rigorous test of the possibility of program recruitment. If the program works in that site, "It could work anywhere."

World-renowned medical hypnotist Milton H. Erickson became a critical case in the field of hypnosis. Erickson was so skillful that he became widely known for "his ability to succeed with 'impossibles'—people who have exhausted the traditional medical, dental, psychotherapeutic, hypnotic and religious avenues for assisting them in their

need, and have not been able to make the changes they desire" (Grinder et al., 1977: 109). If Milton Erickson couldn't help, no one could help. He was able to demonstrate that anyone could be hypnotized.

(S) *Snowball or chain sampling.* This is an approach for locating information-rich key informants or critical cases. The process begins by asking well-situated people: "Who knows a lot about _____? Who should I talk to?" By asking a number of people who else to talk with, the snowball gets bigger and bigger as you accumulate new information-rich cases. In most programs or systems, a few key names or incidents are mentioned repeatedly. Those people or events recommended as valuable by a number of different informants take on special importance. The chain of recommended informants will typically diverge initially as many possible sources are recommended, then converge as a few key names get mentioned over and over.

The Peters and Waterman (1982) study *In Search of Excellence* began with snowball sampling, asking a broad group of knowledgeable people to identify well-run companies. Another excellent and well-known example was Rosabeth Moss Kanter's (1983) study of innovation reported in *The Change Masters*. Her book focused on ten case studies. She began her search for the "best" or "most innovative" companies by getting the views of corporate experts in human resource fields. Nominations for cases to study snowballed from there and then converged into a small number of core cases nominated by a number of different informants.

(9) *Criterion sampling.* The logic of criterion sampling is to review and study all cases that meet some predetermined criterion of importance. This approach is common in quality assurance efforts. For example, the expected range of participation in a mental health outpatient program might be 4 to 26 weeks. All cases that exceed 28 weeks are reviewed and studied to find out what is happening and to make sure the case is being appropriately handled.

Critical incidents can be a source of criterion sampling. For example, all incidents of client abuse in a program may be objects of in-depth evaluation in a quality assurance effort. All former mental health clients who commit suicide within three months of release may constitute a sample for in-depth, qualitative study. In a school setting, all students who are absent more than half the time may merit the in-depth attention of a qualitative case study. The point of criterion sampling is to be sure to understand cases that are likely to be

information-rich because they may reveal major system weaknesses that become targets of opportunity for program or system improvement.

Criterion sampling can add an important qualitative component to a management information system or an ongoing program monitoring system. All cases in the data system that exhibit certain predetermined criterion characteristics are routinely identified for in-depth, qualitative analysis. Criterion sampling also can be applied to identify cases from quantitative questionnaires or tests for in-depth follow-up.

(10) *Theory-based or operational construct sampling.* A more formal basic research version of criterion sampling is theory-based sampling. The researcher samples incidents, slices of life, time periods, or people on the basis of their potential manifestation or representation of important theoretical constructs. The sample becomes, by definition, representative of the phenomenon of interest. An ecological psychologist (see Chapter 3) is interested, for example, in studying the interaction between a person and the environment. Instances of such interaction must be defined based on theoretical premises in order to study examples that represent the phenomenon of interest.

This differs from the more practical sampling in program evaluation. The evaluator doesn't need a theory-based definition of "program" because the entity to be studied is usually legally or financially defined. However, to sample social science phenomena that represent theoretical constructs of interest, one must define the construct to be sampled, such as person-environmental interactions or instances of social deviance, identity crisis, creativity, or power interactions in an organization.

When one is studying people, programs, organizations, or communities, the population of interest can be fairly readily determined. Constructs do not have as clear a frame of reference; neither does time.

The problem with time sampling is that there are no concrete populations of interest, and we are anyway usually restricted to the limited time span over which a study is conducted or to the only slightly longer time span, historically speaking over which the literature on a topic has accumulated. For sampling operational instances of constructs, there is also no concrete target population.... Mostly, therefore, we are forced to select on a purposive basis those particular instances of a construct that past validity studies, conventional practice, individual intuition, or consultation with critically minded persons suggest offer the closest

correspondence to the construct of interest. Alternatively, we can use the same procedures to select multiple operational representations of each construct, chosen because they overlap in representing the critical theoretical components of the construct and because they differ from each other on irrelevant dimensions. This second form of sampling is called multiple operationalism, and it depends more heavily on individual judgment than does the random sampling of persons from a well-designated, target population. Yet *such* judgments, while inevitable, are less well understood than formal sampling methods and are largely ignored by sampling experts. (Cook et al., 1985: 163-64)

"Operational construct" sampling simply means that one samples for study real-world examples (i.e., *operational* examples) of the constructs in which one is interested. Studying a number of such examples is called "multiple operationalism" (Webb et al., 1966).

(11) *Confirming and disconfirming cases.* In the early part of qualitative fieldwork the evaluator is exploring—gathering data and beginning to allow patterns to emerge. Over time the exploratory process gives way to confirmatory fieldwork. This involves testing ideas, confirming the importance and meaning of possible patterns, and checking out the viability of emergent findings with new data and additional cases. This stage of fieldwork requires considerable rigor and integrity on the part of the evaluator in looking for and sampling confirming *as well as* disconfirming cases.

Confirmatory cases are additional examples that fit already emergent patterns; these cases confirm and elaborate the findings, adding richness, depth, and credibility. Disconfirming cases are no less important at this point. These are the examples that don't fit. They are a source of rival interpretations as well as a way of placing boundaries around confirmed findings. They may be "exceptions that prove the rule" or exceptions that disconfirm and alter what appeared to be primary patterns.

The source of questions or ideas to be confirmed or disconfirmed may be from stakeholders or previous scholarly literature rather than the evaluator's fieldwork. An evaluation may in part serve the purpose of confirming or disconfirming stakeholder's or scholars' pre-conceptions, these having been identified during early, conceptual evaluator-stakeholder design discussions or literature reviews.

Thinking about the challenge of finding confirming and disconfirming cases emphasizes the relationship between sampling and

research conclusions. The sample determines what the evaluator will have something to say about—thus the importance of sampling carefully and thoughtfully.

(12) *Opportunistic sampling.* Fieldwork often involves on-the-spot decisions about sampling to take advantage of new opportunities during actual data collection. Unlike experimental designs, qualitative inquiry designs can include new sampling strategies to take advantage of unforeseen opportunities *after* fieldwork has begun. Being open to following wherever the data lead is a primary strength of qualitative strategies in research. This permits the sample to emerge during fieldwork.

When observing, it is not possible to capture everything. It is, therefore, necessary to make decisions about which activities to observe, which people to observe and interview, and what time periods will be selected to collect data. These decisions cannot all be made in advance. The purposeful sampling strategies discussed above often depend on some knowledge of the setting being studied. Opportunistic sampling takes advantage of whatever unfolds as it unfolds.

(13) *Purposeful random sampling.* The fact that a small sample size will be chosen for in-depth qualitative study does not automatically mean that the sampling strategy should not be random. For many audiences, random sampling, even of small samples, will substantially increase the credibility of the results. I recently worked with a program that annually appears before the state legislature and tells "war stories" about client successes, sometimes even including a few stories about failures to provide balance. They decided they wanted to begin collecting evaluation information. Because they were striving for individualized outcomes they rejected the notion of basing the evaluation entirely on a standardized pre-post instrument. They wanted to collect case histories and do in-depth case studies of clients, but they had very limited resources and time to devote to such data collection. In effect, staff at each program site, many of whom serve 200 to 300 families a year, felt that they could only do 10 or 15 detailed, in-depth clinical case histories each year. We systematized the kind of information that would be going into the case histories at each program site and then set up a random procedure for selecting those clients whose case histories would be recorded in depth. Essentially, this program thereby systematized and randomized their collection of "war stories." While they cannot generalize to the entire client

population on the basis of 10 cases from each program site, they will be able to tell legislators that the stories they are reporting were randomly selected *in advance of knowledge of how the outcomes would appear* and that the information collected was comprehensive. The credibility of systematic and randomly selected case examples is considerably greater than the personal, ad hoc selection of cases to report *after the fact*—that is, after outcomes are known.

It is critical to understand, however, that this is a *purposeful random sample*, not a representative random sample. *The purpose of a small random sample is credibility, not representativeness.* A small, purposeful random sample aims to reduce suspicion about why certain cases were selected for study, but such a sample still does not permit statistical generalizations.

(14) *Sampling politically important cases.* Evaluation is inherently and inevitably political to some extent (see Palumbo, 1987; Patton, 1986, 1987b; Turpin, 1989). A variation of the critical case sampling strategy involves selecting (or sometimes avoiding) a politically sensitive site or unit of analysis. For example, a statewide program may have a local site in the district of a state legislator who is particularly influential. By studying carefully the program in that district, evaluation data may be more likely to attract attention and get used. This does not mean that the evaluator then undertakes to make that site look either good or bad, depending on the politics of the moment. This is simply an additional sampling strategy for trying to increase the usefulness and utilization of information where resources permit the study of only a limited number of cases.

The same (broadly speaking) political perspective may inform case sampling in applied or even basic research studies. A political scientist or historian might select the Watergate or Iran-Contra scandals for study not only because of the insights they provide about the American system of government but because of the likely attention such a study would attract. A sociologist's study of a riot or a psychologist's study of a famous suicide would likely involve some attention during sampling to the political importance of the case.

(15) *Convenience sampling.* Finally, there is the strategy of sampling by convenience: doing what's fast and convenient. This is probably the most common sampling strategy—and the least desirable. Too often evaluators using qualitative methods think that, because the sample size they can study is too small to permit generalizations, it doesn't matter how cases are picked, so they might as well pick ones

that are easy to access and inexpensive to study. *While convenience and cost are real considerations, they should be the last factors to be taken into account* after strategically deliberating on how to get the most information of greatest utility from the limited number of cases to be sampled. Purposeful, strategic sampling can yield crucial information about critical cases. *Convenience sampling is neither purposeful nor strategic.*

Information-Rich Cases

Table 5.5 summarizes the 15 purposeful sampling strategies discussed above, plus a 16th approach—combination or mixed purposeful sampling. For example, an extreme group or maximum heterogeneity approach may yield an initial potential sample size that is still larger than the study can handle. The final selection, then, may be made randomly—a combination approach. Thus these approaches are not mutually exclusive. Each approach serves a somewhat different purpose. Because research and evaluations often serve multiple purposes, more than one qualitative sampling strategy may be necessary. In long-term fieldwork all of these strategies maybe used at some point.

These are not the only ways of sampling qualitatively. The underlying principle that is common to all these strategies is selecting information-rich cases. These are cases from which one can learn a great deal about matters of importance. They are cases worthy of in-depth study.

In the process of developing the research design, the evaluator or researcher is trying to consider and anticipate the kinds of arguments that will lend credibility to the study as well as the kinds of arguments that might be used to attack the findings. Reasons for site selections or individual case sampling need to be carefully articulated and made explicit. Moreover, it is important to be open and clear about the study's limitations, including how any particular purposeful sampling strategy may lead to distortion in the findings—that is, to anticipate criticisms that will be made of a particular sampling strategy.

Having weighed the evidence and considered the alternatives, evaluators and primary stakeholders make their sampling decisions, sometimes painfully, but always with the recognition that there are no perfect designs. The sampling strategy must be selected to fit the purpose of the study, the resources available, the questions being

Table 5.5 Sampling Strategies

<i>Type</i>	<i>Purpose</i>
A. Random probability sampling	Representativeness: Sample size a function of population size and desired confidence level.
1. simple random sample	Permits generalization from sample to the population it represents.
2. stratified random and cluster samples	Increases confidence in making generalizations to particular subgroups or areas.
B. Purposeful sampling	Selects information-rich cases for in-depth study. Size and specific cases depend on study purpose.
1. extreme or deviant case sampling	Learning from highly unusual manifestations of the phenomenon of interest, such as outstanding successes/ notable failures, top of the class/ dropouts, exotic events, crises.
2. intensity sampling	Information-rich cases that manifest the phenomenon intensely, but not extremely, such as good students/ poor students, above average/below average.
3. maximum variation sampling-purposefully picking a wide range of variation on dimensions of interest	Documents unique or diverse variations that have emerged in adapting to different conditions. Identifies important common patterns that cut across variations.
4. homogeneous sampling	Focuses, reduces variation, simplifies analysis, facilitates group interviewing.
5. typical case sampling	Illustrates or highlights what is typical, normal, average.
6. stratified purposeful sampling	Illustrates characteristics of particular subgroups of interest; facilitates comparisons.
7. critical case sampling	Permits logical generalization and maximum application of information to other cases because if it's true of this one case it's likely to be true of all other cases.
8. snowball or chain sampling	Identifies cases of interest from people who know people who know people who know what cases are information rich, that is, good examples for study, good interview subjects.

Table 5.5 (continued)

<i>Type</i>	<i>Purpose</i>
9. criterion sampling	Picking all cases that meet some criterion, such as all children abused in a treatment facility. Quality assurance.
10. theory-based or operational construct sampling	Finding manifestations of a theoretical construct of interest so as to elaborate and examine the construct.
11. confirming and disconfirming cases	Elaborating and deepening initial analysis, seeking exceptions, testing variation.
12. opportunistic sampling	Following new leads during fieldwork, taking advantage of the unexpected, flexibility.
13. random purposeful sampling (still small sample size)	Adds credibility to sample when potential purposeful sample is larger than one can handle. Reduces judgment within a purposeful category. (Not for generalizations or representativeness.)
14. sampling politically important cases	Attracts attention to the study (or avoids attracting undesired attention by purposefully eliminating from the sample politically sensitive cases).
15. convenience sampling	Saves time, money, and effort. Poorest rationale; lowest credibility. Yields information-poor cases.
16. combination or mixed purposeful sampling	Triangulation, flexibility, meets multiple interests and needs.

asked, and the constraints being faced. This holds true for sampling strategy as well as sample size.

SAMPLE SIZE

Qualitative inquiry is rife with ambiguities. There are purposeful strategies instead of methodological rules. There are inquiry approaches instead of statistical formulas. Qualitative inquiry seems to work best for people with a high tolerance for ambiguity. (And we're still only discussing design. It gets worse when we get to analysis.)

Nowhere is this ambiguity clearer than in the matter of sample size. I get letters. I get calls. "Is 10 a large enough sample to achieve maximum variation?"

"I started out to interview 20 people for 2 hours each, but I've lost 2 people. Is 18 large enough, or do I have to find 2 more?"

"I want to study just one organization, but interview 20 people in the organization. Is my sample size 1 or 20 or both?"

My universal, certain, and confident reply to these questions is this: "*it depends.*"

There are no rules for sample size in qualitative inquiry. Sample size depends on what you want to know, the purpose of the inquiry, what's at stake, what will be useful, what will have credibility, and what can be done with available time and resources.

Earlier in this chapter, I discussed the trade-off between breadth and depth. With the same fixed resources and limited time, a researcher could study a specific set of experiences for a larger number of people (seeking breadth) or a more open range of experiences for a smaller number of people (seeking depth). In-depth information from a small number of people can be very valuable, especially if the cases are information-rich. Less depth from a larger number of people can be especially helpful in exploring a phenomenon and trying to document diversity or understand variation. I repeat, the size of the sample depends on what you want to find out, why you want to find it out, how the findings will be used, and what resources (including time) you have for the study.

To understand the problem of small samples in qualitative inquiry, it's necessary to place these small samples in the context of probability sampling. A qualitative inquiry sample *only seems small* in comparison with the sample size needed for representativeness when the purpose is generalizing from a sample to the population of which it is a part. Suppose there are 100 people in a program to be evaluated. It would be necessary to randomly sample 80 of those people (80%) to make a generalization at the 95% confidence level. If there are 500 people in the program, 217 people must be sampled (43%) for the same level of confidence. If there are 1,000 people, 278 people must be sampled (28%); and if there are 5,000 people in the population of interest, 357 must be sampled (7%) to achieve a 95% confidence level in the generalization of findings. At the other extreme, if there are only 50 people in the program, 44 must be randomly sampled (88%) to achieve

a 95% level of confidence. (See Fitzgibbon and Morris, 1987: 163, for a table on determining sample size from a given population.)

The logic of purposeful sampling is quite different from the logic of probability sampling. The problem is, however, that the utility and credibility of small purposeful samples are often judged on the basis of the logic, purpose, and recommended sample sizes of probability sampling. What *should* happen is that purposeful samples be judged on the basis of the purpose and rationale of each study and the sampling strategy used to achieve the study's purpose. The sample, like all other aspects of qualitative inquiry, must be judged in context—the same principle that undergirds analysis and presentation of qualitative data. Random probability samples cannot accomplish what in-depth, purposeful samples accomplish, and vice versa.

Piaget contributed a major breakthrough to our understanding of how children think by observing his own two children at length and in great depth. Freud established the field of psychoanalysis based on fewer than ten client cases. Bandler and Grinder (1975a, 1975b) founded neurolinguistic programming (NLP) by studying three renowned and highly effective therapists: Milton Erickson, Fritz Perls, and Virginia Satir. Peters and Waterman (1982) formulated their widely followed eight principles for organizational excellence by studying 62 companies, a very small sample of the thousands of companies one might study.

The validity, meaningfulness, and insights generated from qualitative inquiry have more to do with the information-richness of the cases selected and the observational/analytical capabilities of the researcher than with sample size.

This issue of sample size is a lot like the problem students have when they are assigned an essay to write.

Student: "How long does the paper have to be?"

Instructor: "Long enough to cover the assignment."

Student: "But how many pages?"

Instructor: "Enough pages to do justice to the subject—no more, no less."

Lincoln and Guba (1985: 202) recommend sample selection

to the point of redundancy ... In purposeful sampling the size of the sample is determined by informational considerations. If the purpose

is to maximize information, the sampling is terminated when no new information is forthcoming from new sampled units; thus redundancy is the primary criterion. (emphasis in the original)

This strategy leaves the question of sample size open.

There remains, however, the practical problems of how to negotiate an evaluation budget or how to get a dissertation committee to approve a design if you don't have some idea of sample size. Sampling to the point of redundancy is an ideal, one that works best for basic research, unlimited time lines, and unconstrained resources.

The solution is judgment and negotiation. I recommended that qualitative sampling designs specify *minimum samples* based on expected reasonable coverage of the phenomenon given the purpose of the study and stakeholder interests. One may add to the sample as fieldwork unfolds. One may change the sample if information emerges that indicates the value of a change. The design should be understood to be flexible and emergent. Yet, at the beginning, for planning and budgetary purposes, one specifies a minimum expected sample size and builds a rationale for that minimum, as well as criteria that would alert the researcher to inadequacies in the original sampling approach and/or size.

In the end, sample size adequacy, like all aspects of research, is subject to peer review, consensual validation, and judgment. What is crucial is that the sampling procedures and decisions be fully described, explained, and justified so that information users and peer reviewers have the appropriate context for judging the sample. The researcher or evaluator is absolutely obligated to discuss how the sample affected the findings, the strengths and weaknesses of the sampling procedures, and any other design decisions that are relevant for interpreting and understanding the reported results. Exercising care not to overgeneralize from purposeful samples, while maximizing to the full the advantages of in-depth, purposeful sampling, will do much to alleviate concerns about small sample size.