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 random sampling derive from statistical probability theory. A random and statistically representative sample permits confident generalization from a sample to a larger population. Random sampling also controls for selection bias. The purpose of probability-based random sampling is generalization from the sample to a population and control of selectivity errors.

What would be "bias" in statistical sampling, and therefore a weakness, becomes intended focus in qualitative sampling, and therefore a strength. The logic and power of purposeful sampling lie 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 inquiry. Thus the term purposeful sampling. Studying information-rich cases yields insights and in-depth understanding rather than empirical generalizations. For example, if the purpose of an evaluation is to increase the effectiveness of a program in reaching lower-socioeconomic groups, one can learn a great deal more by studying in depth a small number of carefully selected poor families than by gathering standardized information from a large, statistically representative sample of the whole program. Purposeful sampling focuses on selecting information-rich cases whose study will illuminate the questions under study. Purposeful sampling is sometimes called purposive or judgment sampling: "In judgment sampling, you decide the purpose you want informants (or communities) to serve, and you go out to find some" (Bernard 2000:176). There are several different strategies for purposefully selecting information-rich cases. The logic of each strategy serves a particular purpose.

1. Extreme or deviant case sampling. This strategy involves selecting cases that are
### EXHIBIT 5.5 Examples of Units of Analysis for Case Studies and Comparisons

<table>
<thead>
<tr>
<th>People Focused</th>
<th>Structure Focused</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individuals</td>
<td>Projects</td>
</tr>
<tr>
<td>Small, informal groups (friends, gangs)</td>
<td>Programs</td>
</tr>
<tr>
<td>Families</td>
<td>Organizations</td>
</tr>
<tr>
<td></td>
<td>Units in organizations</td>
</tr>
</tbody>
</table>

**Perspective/Worldview Based**

- People who share a culture
- People who share a common experience or perspective, for example, dropouts, graduates, leaders, parents, Internet listserv participants, survivors

**Geography Focused**

- Neighborhoods
- Cities
- States
- Countries

- Villages
- Farms
- Regions
- Markets

**Activity Focused**

- Critical incidents
- Celebrations
- Quality assurance violations

- Time periods
- Crises
- Events

**Time Based**

- Particular days, weeks, or months
- Winter break
- Ramadan
- Full moons
- A political term of office

- Vacations
- Rainy season
- Dry season
- School term
- An election period

NOTE: These are not mutually exclusive categories.
influential book *Within Our Reach*. Stephen Covey’s (1990) best-selling book *The 7 Habits of Highly Effective People* is based on a purposeful, extreme group sampling strategy. Studies of leadership have long focused on identifying the characteristics of highly successful leaders, as in Collins’s (2001) case studies of 11 corporate executives in whom “extreme personal humility blends paradoxically with intense professional will” (p. 67), what he calls “Level 5 leaders,” the highest level in his model. In the early days of AIDS research when HIV infections almost always resulted in death, a small number of cases of people infected with HIV who did not develop AIDS became crucial outlier cases that provided important insights into directions researchers should take in combating AIDS.

Sometimes cases of dramatic failure offer powerful lessons. The legendary UCLA basketball coach John Wooden won 10 national championships from 1964 through 1975, an unparalleled sports achievement. But the game he remembered the most and said he learned the most from was UCLA’s 1974 overtime loss to North Carolina State in the semifinals (quoted in the *Los Angeles Times*, December 21, 2000). Wooden’s focus on that game—that extreme case—illustrates the learning psychology of extreme group purposeful sampling.

This is also the sampling psychology behind Jim Paul’s (1994) book *What I Learned Losing a Million Dollars*. A former governor of the Chicago Mercantile Exchange, he made thousands of trades in many commodities over a long and distinguished career, but what he reports learning the most from was a highly unusual combination of mistakes in which he lost more than $1 million in a few weeks of trading soy beans—an extreme but illuminative case. He reports that he ultimately learned to be a winner by carefully studying and learning from losing.

An excellent applied research example 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 is what renders them so powerful. Browne’s book is an exemplar of qualitative inquiry using purposeful sampling for applied research.

In evaluation, 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 “OK.” 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 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 illuminative 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 even necessary to random programs or excellent researchers and intended study think through what the most from are that are selected for study.

In a single program, may apply. Instead of a representative sample of people evaluator may focus on
SAMPLING THE BEST

Evaluating the Caribbean Agricultural Extension Project posed a challenge: how to provide funders with decision-relevant information about the long-term potential of agricultural extension in eight English-speaking Caribbean countries. The first phase of the project involved needs assessment, planning, and capacity building. These processes laid the foundation for training agricultural extension agents who would, it was hoped, help improve the productivity and profitability of small farms. However, a critical funding decision about the potential of the project had to be made at the end of the first phase before actual training of extension agents had begun and prior to any impact on farmers. No data existed on the effectiveness of extension that could be used to calculate potential effectiveness. Funders were asking for concrete estimates of future impact, not just some hoped-for productivity increase pulled out of the air. The solution was to study a purposeful sample of "the best."

Each agricultural extension service established a process for identifying and recognizing its own "outstanding agricultural extension agent." Those agents were each asked to identify five farm families whose farm productivity the agent believed he or she had increased. Independent case studies were done on these 40 farm families (5 families in each of the 8 countries).

This sample was purposefully "biased" to establish case-based goals for increased extension effectiveness while also showing the diversity of small farm situations and the variety of extension agent practices. It could be expected that the typical extension agent would have somewhat less impact on farmers than these outstanding agents. However, by gathering data about the impacts of the "best," it was possible to provide project funders with concrete examples of what might be accomplished over time if more extension agents were trained in the practices of the "best." These data allowed potential second-phase funders to engage the question: Given the impacts of those extension agents identified as the best, is it worth funding a training program aimed at creating more extension agents following those "best" practices? Without these concrete cases to examine, the funding decision would have been based on abstract discussion and speculative guesses about extension agent activities and impacts. Having real cases to examine made the resulting discussions concrete, focused, and data-based. The second phase was funded.

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 an evaluation of the Caribbean Agricultural Extension Project, we did case studies of the "outstanding extension agent" selected by peers in each of eight Caribbean countries to help the program develop curriculum and standards for improving extension practice. The sample was purposefully "biased," not to make the program look good, but rather to
learn from those who were exemplars of good practice. In many instances, more can be learned from intensively studying exemplary (information-rich) 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. In statistical terms, extreme case sampling focuses on outliers (the endpoints of the bell-shaped curve normal distribution) that are often ignored in aggregate data reporting.

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 greatly 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.

In essence, the logic of extreme group sampling is that extreme cases may be information-rich cases precisely because, by being unusual, they can illuminate both the unusual and the typical. In proposing an extreme group sample, as in all purposeful sampling designs, the researcher has an obligation to present the rationale and expected benefits of this strategy as well as to note its weakness (lack of generalizability).

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 highly unusual cases.

Heuristic research (Chapter 3) uses intensity sampling. Heuristic research draws explicitly on the intense personal experiences of the researcher, for example, experiences with loneliness or jealousy. Co-researchers who have experienced these phenomena intensely also participate in the study. The heuristic researcher is not typically seeking pathologically 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 co-researchers 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 strategy can be applied in a program evaluation. Extreme successes or unusual failures may be discredited as being too extreme or unusual to yield useful 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 or evaluator must do some exploratory work to determine the nature of the variation in the situation under study, then sample intense examples of the phenomenon of interest.

3. *Maximum variation (heterogeneity) sampling.* This strategy for purposeful sampling aims at capturing and describing the central themes that constitute an important variation. For small projects, heterogeneity can best be handled by individual cases and their interactions with each other. The maxim: no sample is too small to study. Variation is a strength by approach, not a weakness.

Any common project problem has great variation and richness of value in capturing this variation centrally, shared data about this phenomenon.

How does one develop a small sample? One aspect of diverse characteristics is selecting the sample. Standardization program has projected the state, as some in rural areas, and some in urbanization lacks sufficient data. To select enough projects across the state. The few sites from each area so that the geographic area is represented in the evaluation. That is, the description would describe what would happen at any part of the state, it would also lend itself to projects across sites. Any such project importance precisely out of great variation in existing community-based and state efforts statewide using heterogeneity sampling strategy. And they have matrix sample of 10 communities, each community was from every other characteristic as size, for example (e.g., strong mayor/weak mayor, diversity, strength of leadership, demographics, and region). These communities stood out across these characteristics. This importance of a local network or the people who made things happen.

In a study of the Model Fellowship Program,
central themes that cut across a great deal of 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 dimensions of a setting or phenomenon.

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 study a few sites from each area and at least be sure that the geographical variation among sites is represented in the study. While the evaluation would describe the uniqueness of each site, it would also look for common themes across sites. Any such themes take on added importance precisely because they emerge out of great variation. For example, in studying community-based energy conservation efforts statewide using a maximum heterogeneity sampling strategy, I constructed a matrix sample of 10 communities in which each community was as different as possible from every other community on such characteristics as size, form of local government (e.g., strong mayor/weak mayor), ethnic diversity, strength of the economy, demographics, and region. In the analysis, what stood out across these diverse cases was the importance of a local, committed cadre of people who made things happen.

In a study of the MacArthur Foundation Fellowship Program, the design focused on case studies of individual fellowship recipients. Over 20 years of awards, more than 600 people had received fellowships. We had sufficient resources to do only 40 case studies. We maximized sample variation by creating a matrix in which each person in the sample was as different as possible from others using dimensions such as nature of work, stage in career, public visibility, institutional affiliation, age, gender, ethnicity, geographic location, mobility, health status, nationality, and field of endeavor. The thematic patterns of achievement that emerged from this diversity allowed us to construct a model to illuminate the primary dimensions of and factors in the award’s impact. A theme song emerged from all the scattered noise. That’s the power of maximum variation (heterogeneity) sampling.

Thus, 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 uniquenesses, and (2) important shared patterns that cut across cases and derive their significance from having emerged out of heterogeneity. Both are important findings in qualitative inquiry.

4. Homogeneous samples. In direct contrast to maximum variation sampling is the strategy of picking a small, homogeneous sample, the purpose of which 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 based typically on homogeneous groups. Focus groups involve 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 issues that affect them.

5. Typical case sampling. In describing a culture or program to people not familiar with the setting studied, 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 who and what are typical. Typical cases can also be selected using survey data, a demographic analysis of averages, or other statistical data that provide a normal distribution of characteristics from which to identify “average-like” cases. 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 setting—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, the processes and effects described for the typical program may be used to provide a frame of reference for case studies of “poor” 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 that kind of village.

In evaluation and policy research, the interests of decision makers will shape the sampling strategy. I remember an evaluation in which the key decision makers had made their peace with the fact that there will always be some poor programs and some excellent programs, but the programs they really wanted more information about were what they called “those run-of-the-mill programs that are so hard to get a handle on precisely because they are so ordinary and don’t stand out in any definitive way.” Given that framing, we employed typical case sampling. It is important, when using this strategy, to attempt to get broad consensus about which cases are typical—and what criteria are being used to define typicality.

6. Critical case sampling. 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.” 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 for a critical case.

Physics provides an example of a critical case. In 1900, Lord Kelvin said he wanted to find the age of the earth. His calculations would fall. Rather than cast down objects of different materials to deduce the age of the earth, he showed that if the earth were a critical case—the fact that it expanded faster as it heated up—there was both useful and accurate way to deduce the earth’s age was a convincing technique.

Critical cases can be used in evaluation and policy analysis in looking for and shaping national policy. Suppose national policy is needed for the communities involved in the evaluation to consider how their local conditions can be used to make sense of the overall strategy. The critical case is the program of well-known local officials who can’t understand their actions in the context of the national regulations incomprehensible. Suppose the community consists of clients at a lower levels of education. “Critical case” analysis might consider the critical case a program of education for the less-educated, anyone.

Identification of critical cases requires the recognition of the key issue for a critical case. For each case, he might come from a particular program location. If the program is expected to produce the most rigorous conclusions about the program is expected to provide the most rigorous results.
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 comparison to all objects. His finding was both useful and credible because the feather was a convincing critical case.

Critical cases can be found in social science and evaluation 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. For example, 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 program recruitment. If the program works in that site, it could work anywhere. That makes the critical case an especially information-rich exemplar, therefore worthy of study as the centerpiece in a small or "N of 1" sample.

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, DeLozier, and Bandler 1977:109). If Milton Erickson couldn't hypnotize a person, no one could. He was able to demonstrate that, under his definition of hypnosis, anyone could be hypnotized.

7. 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 ____? Whom 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 would 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. Rosabeth Moss Kanter's (1983) study of innovation reported in The Change Masters focused on 10 core case studies of the "most
innovative” companies. She began by asking corporate experts for candidate companies to study. Nominations snowballed as she broadened her inquiry and then converged into a small number of core cases nominated by a number of different expert informants.

8. **Criterion sampling.** The logic of criterion sampling is to review and study all cases that meet some predetermined criterion of importance, a strategy 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 to find out why the expected range was exceeded and to make sure the case was being appropriately handled. Or a quality assurance standard may be that all patients entering a hospital emergency room, who are not in a life-threatening situation, receive care within 2 hours. Cases that exceed this standard are reviewed.

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 25% or more of the time may merit the in-depth attention of a 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 used to identify cases from standardized questionnaires for in-depth follow-up, for example, all respondents who report having experienced ongoing workplace discrimination. (This strategy can only be used where respondents have willingly supplied contact information.)

9. **Theory-based sampling, operational construct sampling, and theoretical sampling.** A more conceptually oriented 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. Buckholt (2001) studied people who met theory-derived criteria for being “resilient” in a study of resilience among adult abuse survivors. The sample becomes, by definition and selection, representative of the phenomenon of interest.

When one is studying people, programs, organizations, or communities, the population of interest can be fairly readily determined. Constructs, however, do not have as clear a frame of reference:

For sampling operational instances of constructs, there is 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 are not on irrelevant or unconstructive cases. The sampling is clearly judgmental, and it depends on the judgment that persons from a larger population. Yet, such people are less well understood by other methods and are not necessarily experts. (Cook 1963-64)

Operational constructs means that one can identify real-world examples of the substantive domain of the constructs being examined.

Studying a number of innovations, for example, or what is called “multiple cases,” requires that early adopters and late adopters be different in significant respects. Doing complex sampling, theory-based sampling is often necessarily judgmental, because the population of early adopters may not be the same as the population of late adopters, and the sampling is not necessarily random.

Theoretical sampling, theorists define as “the systematic collection of the emerging concepts that are guiding the emerging concepts that are guiding the emerging concepts that are guiding the emerging concepts that are guiding the emerging concepts that are guiding the emerging concepts.” In grounded theory sampling supports the constant comparative method of analysis. Theoretical sampling in grounded theory sampling permits the constant comparison of the variations in, and the meanings of a conce...
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). For example, classic diffusion of innovations theory (Rogers 1962) predicts that early adopters of some innovation will be different in significant ways from later adopters. Doing cases studies on early and late adopters, then, would be an example of theory-based sampling. Such samples are often necessarily purposefully selected because the population of all early and late adopters may not be known, so random sampling is not an option.

Theoretical sampling is what grounded theorists define as "sampling on the basis of the emerging concepts, with the aim being to explore the dimensional range or varied conditions along which the properties of concepts vary" (Strauss and Corbin 1998: 73). In grounded theory, theoretical sampling supports the constant comparative method of analysis. That is, one does theoretical sampling in grounded theory in order to use the constant comparative method of analysis. The two go hand in glove, connecting design and analysis. Theoretical sampling permits elucidation and refinement of the variations in, manifestations of, and meanings of a concept as it is found in the
data gathered during fieldwork. The constant comparative method involves systematically examining and refining variations in emergent and grounded concepts. Variations in the concept must be sampled to rigorously compare and contrast those variations. (See Chapters 3 and 8 for more detailed discussions of grounded theory.)

10. Confirming and disconfirming cases. In the early part of qualitative fieldwork, the evaluator is exploring—gathering data and watching for 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 stakeholders’ or scholars’ hypotheses, 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.

11. Stratified purposeful sampling. Stratified samples are samples within samples. A stratified random sample, for example, might stratify by socioeconomic status within a larger population so as to make generalizations and statistically valid comparisons by social class as well as to generalize to the total population.

Purposeful samples can also be stratified and nested by combining types of purposeful sampling. So, for example, one might combine typical case sampling with maximum heterogeneity sampling by taking a stratified purposeful sample of above average, average, and below average cases. This represents less than a full maximum variation sample, but more than simple typical case sampling. 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.

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

During fieldwork, it is impossible to observe everything. Decisions must be made about what activities to observe, which people to observe and interview, and when to collect data. These decisions cannot all be made in advance. The purposeful sampling strategies discussed above provide direction for sampling but often depend on some knowledge of the setting being studied. Opportunistic, emergent sampling takes advantage of whatever unfolds as it unfolds.

In Chapter 2, I identified emergent flexible designs as one of the core strategic themes of qualitative inquiry and cited as an exemplar the anthropologist Brackette F. Williams and her fieldwork on how Americans view violence in America:

I do impromptu interviews. I don't have some target number of interviews in mind or predetermined questions. It depends on the person and the situation. Airports, for example, are a good place for impromptu interviews with people. So sometimes, instead of using airport time to write, I interview people about the death penalty or about killing or about death in their life. It's called opportunity sampling. I'm following where the data take me, where my questions take me. (personal interview)

Few qualitative studies are as fully emergent and open-ended as the fieldwork of Williams. Her approach exemplifies emergent opportunity sampling.

13. Purposeful random sampling. A purposeful random sampling strategy does not automatically eliminate any possibility for random selection of cases. For many audiences, random sampling, even of small samples, will substantially increase the credibility of the results. I recently learned of a study that annually appraises the literature and tells of successes and struggles, including a few stories of providing balance. To keep their reports, the researchers began to collect more systematically, striving for individuals rejected the notion that they should rely entirely on a single client's account. They wanted to talk to many and do in-depth cases. Like they had very limited amounts of time to devote to such data collection, so at each program site they would collect data from 200 to 300 families a year. They would collect data on only do 10 or 15 different cases, and they would spend 3 to 4 hours collecting those cases. They would then write each case history each family and then set up a random sample in around each case. The kind of information collected would be recorded systematically and then set up a random sample. Information was collected in advance of knowing what outcomes would arise. The information collected was then used to assess the credibility of systematic selected case examples in the larger than the personal, ad hoc selected and reported after outcomes are known.

It is critical to understand that this is a purposeful random sampling. A representative random sample is a small random sample that represents. A random sample aims to
results. I recently worked with a program that annually appears before the state legislature and tells "war stories" about client successes and struggles, sometimes even including a few stories about failures to provide balance. To enhance the credibility of their reports, the director and staff decided to begin collecting evaluation information more systematically. 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, thereby systematizing and randomizing 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 selected and reported 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 (see Turpin 1989; Palumbo 1987; Patton 1987b). A variation on the critical case 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. That would clearly be unethical. Rather, sampling politically important cases is simply a strategy for trying to increase the usefulness and relevance of information where resources permit the study of only a limited number of cases.

The same political perspective (broadly speaking) may inform case sampling in applied or even basic research studies. A political scientist or historian might select the election year 2000 Florida vote-counting case, the Clinton impeachment effort, Nixon's Watergate crisis, or Reagan's Iran-Contra scandal 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 public and 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 will be 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

Exhibit 5.6 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 may be used at some point.

The underlying principle that is common to all these strategies is selecting information-rich cases—cases from which one can learn a great deal about matters of importance and therefore 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 a study’s limitations, that is, to anticipate and address criticisms that may be made of a particular sampling strategy, especially from people who think that the only high-quality samples are random ones.

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 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. I get e-mails.

Is 10 a large enough sample to achieve maximum variation?

I started out to interview 20 people for two hours each, but I’ve lost 2 people. Is 18 large enough, or do I have to find 2 more?
### EXHIBIT 5.6 Sampling Strategies

<table>
<thead>
<tr>
<th>Type</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Random probability sampling</strong></td>
<td>Representativeness: Sample size a function of population size and desired confidence level.</td>
</tr>
<tr>
<td>1. Simple random sample</td>
<td>Permit generalization from sample to the population it represents.</td>
</tr>
<tr>
<td>2. Stratified random and cluster samples</td>
<td>Increase confidence in making generalizations to particular subgroups.</td>
</tr>
<tr>
<td><strong>Purposeful sampling</strong></td>
<td>Select information-rich cases strategically and purposefully; specific type and number of cases selected depends on study purpose and resources.</td>
</tr>
<tr>
<td>1. Extreme or deviant case (outlier) sampling</td>
<td>Learning from unusual manifestations of the phenomenon of interest, for example, outstanding successes/notable failures; top of the class/dropouts; exotic events; crises.</td>
</tr>
<tr>
<td>2. Intensity sampling</td>
<td>Information-rich cases that manifest the phenomenon intensely, but not extremely, for example, good students/poor students; above average/below average.</td>
</tr>
<tr>
<td>3. Maximum variation sampling—purposefully picking a wide range of cases to get variation on dimensions of interest</td>
<td>Document unique or diverse variations that have emerged in adapting to different conditions. Identify important common patterns that cut across variations (cut through the noise of variation).</td>
</tr>
<tr>
<td>4. Homogeneous sampling</td>
<td>Focus; reduce variation; simplify analysis; facilitate group interviewing.</td>
</tr>
<tr>
<td>5. Typical case sampling</td>
<td>Illustrate or highlight what is typical, normal, average.</td>
</tr>
<tr>
<td>6. Critical case sampling</td>
<td>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.</td>
</tr>
<tr>
<td>7. Snowball or chain sampling</td>
<td>Identify cases of interest from sampling people who know people who know people who know what cases are information rich, that is, good examples for study, good interview participants.</td>
</tr>
<tr>
<td>8. Criterion sampling</td>
<td>Picking all cases that meet some criterion, for example, all children abused in a treatment facility. Quality assurance.</td>
</tr>
<tr>
<td>9. Theory-based sampling, operational construct sampling, or theoretical sampling</td>
<td>Finding manifestations of a theoretical construct of interest so as to elaborate and examine the construct and its variations.</td>
</tr>
</tbody>
</table>

(continued)
gram 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 Fitz-Gibbon and Morris [1987:163] for a table on determining sample size from a given population.)

The logic of purposeful sampling is quite different. 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. Instead, purposeful samples should be judged according to the purpose and rationale of the study: Does the sampling strategy support 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 originally on fewer than 10 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. Sands (2000) did a fine dissertation studying a single school principal, describing the leadership of a female leader who entered a challenging school situation and brought about constructive change.

Clair Claiborne Parker’s [2001] single case study of her daughter’s autism reports 40 years of data on every stage of her development, language use, emotions, capacities, barriers, obsessions, communication patterns, emergent artistry, and challenges overcome and challenges not overcome. Parker and her husband made systematic observations throughout the years. Eminent medical anthropologist Oliver Sacks reviewed the data and determined in his preface to the book that more data are available on the woman in this extraordinary case study than on any other autistic human being who has ever lived. Here, then, is the epitome of N = 1, in-depth inquiry.

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) 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” (p. 202).

This strategy leaves the question of sample size open, another example of the emergent nature of qualitative inquiry. There remains, however, the practical problem of how to negotiate an evaluation budget or 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 timelines, and unconstrained resources.

The solution is judgment and negotiation. I recommend 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 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.

Emergent Designs and Protection of Human Subjects

Emergent designs pose special problems for institutional review boards (IRBs) charged with approving research designs to ensure protection of human subjects. Such boards typically want to know, in advance of fieldwork, who will be interviewed and the precise questions that will be asked. If the topic is fairly innocuous and the general line of questioning relatively unobtrusive, an IRB may be willing to approve the framework of an emergent design with sample questions included, but without full sample specification and a formal interview instrument.

Another approach is to ask for approval in stages. This means initially asking for approval for the general framework of the inquiry and specifically for the first exploratory stage of fieldwork, including procedures for assuring confidentiality and informed consent, then returning periodically (e.g., quarterly or annually) to update the design and its approval. This is cumbersome for both the researcher and the IRB, but it is a way of meeting IRB mandates and still implementing an emergent design. This staged-approval approach can also be used when the evaluator is developing the design jointly with program staff and/or partici-