SELECTING A SAMPLE OF NONGENERALIZABLE
CASES FOR REVIEW IN GAO ENGAGEMENTS

N**ote**: This guidance is designed to ensure that GAO policies on evidence and generally accepted government auditing standards are met. The guidance conforms to the generally accepted principles and practices of the appropriate disciplines. Statements that particular actions "should" be taken are practices that are expected to be followed, unless there are good reasons for not doing so. Before deviating from a practice expressed as a "should" statement, staff members must consult with an appropriate staff member in Applied Research and Methods (ARM) or a team specialist and must document the consultation.

Abstract: This paper identifies ways of selecting nongeneralizable sites and cases for in-depth study—for example, choosing case studies, identifying sites to visit, selecting documents to review, or selecting people to interview. This sampling approach is known as nongeneralizable, or purposeful or judgmental, sampling.[1] The approach seeks to systematically identify cases that will be useful for answering researchable questions. The paper includes guidance on (A) selecting and using a nongeneralizable sample, (B) documenting selection decisions in workpapers, (C) reporting data collected by using nongeneralizable sampling, and (D) describing selection decisions in GAO reports. Appendix I contains tables describing strategies for making selections; appendix II provides objectives, scope, and methodology (OSM) language from GAO reports that used nongeneralizable samples; and appendix III is a workpaper template that can be used to help document decisions.

A. NONGENERALIZABLE SAMPLE SELECTION DECISION STEPS
When selecting a sample of sites or cases, GAO staff should take the following five steps:
1. ensure that a nongeneralizable sample is appropriate for your purposes,
2. determine the appropriate sampling strategy and criteria,
3. specify cases from which to collect data,
4. determine the number of cases to select, and
5. enhance the validity and reliability of your evidence.

You should also document the factors you considered in making these decisions. Each step is discussed below.

Step 1. Ensure That a Nongeneralizable Sample Is Appropriate for Your Purposes

Carefully define the purpose of your sample and use that to guide your decision about whether to use a nongeneralizable or generalizable sample.[2] You should consider issues such as whether you want information from the sample to provide context sophistication or illustrative examples or whether you want it to serve as the primary source of evidence in answering the objective. Similarly, you should ask yourself how information from the sample will be reported. As its

---

[1] The financial auditing literature refers to nongeneralizable sampling approaches as "selections" (reserving "samples" for generalizable sampling approaches). If you are conducting a financial audit, you may use this terminology instead.

[2] For additional guidance on determining an appropriate sampling approach, see the ARM guidance, Using Probability, Nonprobability, and Certainty Samples.

name indicates, findings from a nongeneralizable sample may not be extrapolated beyond your sample.

If your objective focuses on a specific group within a population, or is to describe aspects of an issue, understand the context of a problem, or provide anecdotes to illustrate a finding, then a nongeneralizable sample would be appropriate. GAO regularly uses nongeneralizable samples to obtain in-depth knowledge about a small number of cases.

If, in contrast, your objective is to report generalizations about a population, such as the percentage of an agency's officials who received certain training, or the total dollar value of transactions in error in an agency's system, then a generalizable or probability sample would be appropriate.

You may have to use a nongeneralizable sample if you would like to select a generalizable sample but cannot because, for example, you lack reliable data about the population of interest. In this situation, a sound nongeneralizable sampling strategy could allow you to obtain useful information. Sometimes, the available data show that a small nongeneralizable sample represents a large proportion of a population. For example, if 20 grantees of a federal program accounted for 75 percent of the program's expenditures, you might decide that your researchable questions could be answered with definitive statements about the grantees that represented three-quarters of expenditures.

Finally, GAO commonly uses a combination of generalizable and nongeneralizable samples to answer research objectives. For example, you might select a generalizable sample of agency officials to survey, as well as a small nongeneralizable sample of sites to visit for more in-depth information.

Step 2. Determine the Appropriate Nongeneralizable Sampling Strategy and Criteria

Once you have determined that a nongeneralizable sample suits your engagement, you will need to determine how to select cases from which to collect data. Your strategy for sample selection should be directly linked to your researchable objective and the purpose for which you have chosen a nongeneralizable sample. When selecting a strategy, think about the characteristics of the cases that will best enable you to answer your researchable objectives. For example, you may be interested in the most recent or the largest, the worst or the best. Various strategies for selecting samples are described in appendix I, "Nongeneralizable Sampling Strategies."

You will also need to develop relevant criteria to use in conjunction with the sampling strategy. The criteria should be directly linked to the researchable questions. For example, you may have elected a stratified purposeful sampling strategy. If your research objective concerned the efficacy of widget screeners and you suspected that screener model and location were relevant to failure rates, you might select criteria that allow you to cover a range of screeners with varying failure rates, screeners of different makes or models, and screeners across a variety of locations.

It is often possible to obtain or develop a list of the population you will select cases from that is based on characteristics important to your researchable question. Sometimes, however, such a

list is not available or cannot be created. Appendix I describes sampling strategies you can use when you have a list, part of a list, or no list.

Step 3. Specify Cases from Which to Collect Data

In step 3, you want to ask, "Who or what is best suited to answer the research question?" The "who or what" to collect data from is a "case" or "unit of analysis." In nongeneralizable sampling, the unit of analysis may be any of the following:

a.   **a person:** federal employee, program recipient, program director.

b.   **an organization:** federal agencies, local school districts, government contractors.

c.   **a place:** national parks, highways, former military bases, foreign countries.

d.   **a process:** hiring federal air marshals, identifying unexploded ordinance, awarding federal contracts, managing agency information technology systems.

e.   **an event:** major wildfires, UN peacekeeping missions.

f.   **a program:** federal environmental regulations, aid to Afghanistan.

g.   **a document or file:** elementary school curriculum, written evaluation of military unit readiness, corporate financial report.

h.   **other:** endangered species, computer systems, prescription drugs.

Units or subunits of analysis can be nested within larger units of analysis. This may require a sampling strategy for each subunit—for each stage of sampling. For example, in sampling a group of charter schools, you might also be interested in sampling a set of classrooms within each sampled school and a group of students within each sampled classroom. Having a clear sampling strategy for selecting cases at each of these stages will enhance the validity and reliability of your study.

Step 4. Determine the Number of Cases to Select

Determining the number of cases to select often requires balancing a sample that is large enough to provide a sufficiently comprehensive understanding of the issues with one that is small enough to study within your time and resource constraints. The number of cases you must review depends on (1) what you want to report, (2) how the findings will be used, (3) what is needed to ensure credibility, and (4) what can be accomplished with available time and resources. For example, you may want to examine a specific set of experiences with a larger number of cases or explore an open range of experiences with a smaller number. Less depth in a review of a large number of cases can be helpful in exploring a specific phenomenon and trying to document diversity or understand variation, whereas greater depth of review of a smaller number of cases may provide you with better understanding of the specific phenomenon. A nongeneralizable

sample is well designed if it meets the purpose and rationale of the study.[3]

A sample is large enough when no new information about your topic of interest is provided from additional sampled cases. This is known as "saturation" or "redundancy." Although saturation is the ideal, it may not be practical to leave sampling open-ended for engagement planning purposes. Early in the design phase, teams should determine the minimum number of cases to review, given the purpose of the sample, the use that is planned for the findings, and the credibility of the evidence, and they should include these details in the design matrix. (Documentation is discussed below.) Michael Patton provides additional information on case selection theory.[4]

Samples often have to be selected when population lists are not complete. For example, sometimes the available lists include only the top dollar recipients of a federal program, not recipients of smaller amounts. In other instances, teams have to create lists from agency data, publicly available information, or other sources. Although it is recommended that teams be as thorough as possible, it may not be possible to construct a list that contains all cases or all descriptive information about cases.

Some of the sampling strategies in appendix I can be applied to incomplete or partial lists; others provide options when no reasonable list can be created. Another approach is to identify and carefully screen potential cases to ensure that they meet the sampling criteria. It is advised that an ARM stakeholder be involved in decisions about the sufficiency of available lists, steps to create lists, and alternative selection techniques.

Step 5. Enhance the Validity and Reliability of Your Evidence

Finally, a well-designed sampling strategy can improve the validity and reliability of your evidence while a poorly designed sampling strategy may lead to insufficient evidence. Validity (as defined in the Yellow Book) refers to the extent to which evidence is a meaningful or reasonable basis for measuring what is being evaluated. In other words, validity refers to the extent to which evidence represents what it is purported to represent. Reliability, which includes the concept of sufficient, appropriate evidence, is integral to an audit. Appropriateness is the measure of the quality of evidence that encompasses its relevance, validity, and reliability in providing support for findings and conclusions related to the audit objectives.

B. DOCUMENTING SELECTION DECISIONS IN WORKPAPERS

It is important to document well how your team implemented a nongeneralizable sample, in order to give credibility to the study and demonstrate that a defensible methodology was used. Reasons for case selections and limitations of the approach should be explicit and well articulated.[5] In the workpapers, you should include items related to case selection that do the

---

[3] See Michael Quinn Patton, *Qualitative Research and Evaluation*, 3rd ed. (Thousand Oaks, Calif.: Sage Publications, 2002), p. 245

[4] Patton, *Qualitative Research and Evaluation*, 3rd ed., pp. 230–46.

[5] Patton, *Qualitative Research and Evaluation*, 3rd ed., p. 242.

following three things.

Item 1. Describe the Source of the List from Which Cases Were Selected

In addition to a description of the source of the list from which cases were selected, the workpapers should include information about list totals, overall and by strata, if applicable. Lists (population information) can be developed by teams or provided by agencies or other sources. Workpapers should include information on (1) the sources— interviews, documents, data systems—used to develop the lists for case selection and (2) evidence that the sources and lists meet GAO's evidence standards. Lists developed from computer processed data should include data reliability assessment documentation.[6]

Item 2. Document Factors in Deciding Selection Strategies and Criteria for Case Selection Decisions

This should include the actual strategies and criteria used for case selection and how they relate to the research questions.

Item 3. Describe the Strengths and Weaknesses of the Evidence

Include in the workpapers a description of the strengths and weaknesses of the evidence, including the strengths and weaknesses of the lists you used for case selection, the sampling strategy, the validity of your selection criteria as they relate to the research objective, and the reliability of the end product of your review, given the strategies you used to ensure accuracy and *consistency* in data collection. Appendix III is an example of a workpaper that you could use to document your decisions.

C. REPORTING DATA COLLECTED FROM NONGENERALIZABLE SAMPLES IN GAO REPORTS

Information from nongeneralizable samples has various uses, depending on your purpose in drawing the sample, the strategies you used to select the sample, and how the information will be used as an evidence source—background, one of multiple sources of evidence, sole source for findings or recommendations.

To ensure high-quality evidence, we often rely on multiple evidence sources and varying methods for acquiring them. Consequently, information from nongeneralizable samples may be used in several ways in the same report and in conjunction with information from a mixture of other sources. Four ways in which we use information from nongeneralizable samples in GAO reports are context sophistication, illustrative example, comparative case, and threshold.

---

[6] See GAO, *Assessing the Reliability of Computer-Processed Data,* GAO-09-365G (Washington, D.C.: February 2009).

Context Sophistication

Nongeneralizable data can provide in-depth information on particular issues or problems, affording teams a more sophisticated understanding of the issues in the audit. Context sophistication gained from nongeneralizable sample information can improve a team's ability to evaluate evidence related to the audit or to accurately describe complex issues in the report. For this use, data might not be made explicit in the report, or in the background section it might provide context for understanding the issues.

Illustrative Example

We often want to report in-depth information about a particular problem, case, or location to illustrate the issue we are trying to describe. Illustrative examples can bolster an argument, demonstrate consequences, and provide practical significance to issues that might, without context, appear to be inconsequential. For example, you learn that Agency X's enforcement policies are decidedly lax, and you want to illustrate some practical consequences of lax enforcement. You visit selected sites that are subject to the agency's enforcement efforts in order to observe compliance issues. You use the information you gather from your visits to describe some of the practical consequences of the agency's lax enforcement efforts.

Comparative Case

Although nongeneralizable cases cannot be extrapolated, they can be used for making comparisons. For example, if you have an established ideal, or best practice case, you can use nongeneralizable sample information to compare to that ideal. You might know that one U.S. Navy platform represents best practice in a particular manning strategy. You collect data from your best practice platform and other selected naval platforms and compare results to identify problem areas. Although you cannot assume that the problems you identified exist across all naval platforms, they nevertheless represent areas for improvement and might be used to develop recommendations.

Threshold

Sometimes, we need only establish that one case or a few cases have certain characteristics in order to show that a significant threshold has been reached. For example, if all nuclear installations are required to have certain security characteristics, and in our site visits to a handful we found that not all installations had such characteristics, we might use this as evidence to support the need for oversight improvements. A second example might be a proposed change to a tax form shown to a group of paid tax preparers who, in discussion, indicated that they did not understand the change. You could reasonably conclude that if paid tax preparers do not understand the change, the average citizen is unlikely to understand it.

D. DESCRIBING SELECTION DECISIONS IN GAO REPORTS

The rationale for case selection should be described in the report's OSM section. Whether the description is in a technical appendix, the introductory OSM, or both, it should be complete enough and technically accurate so as to demonstrate to readers why the particular cases were selected. The description should contain the following four elements.

1.   A Description of the Sampling Strategy and Criteria Used to Make Case Selections

This description should show how the sampling strategy and the criteria used to make case selections relate to the research questions. The description should also include, to the extent it is appropriate, mentions of methods, information sources, and alternative sampling approaches that were considered but not used.

2.   A Clear Definition of the Target Population and Study Population

Include information on the size of the target population and relevant strata, if applicable—for example, "we reviewed 10 of the 20 installations where personnel records are retained." The target population is the larger group from which the study population is drawn, about which the researcher would like to make statements. The study population consists of members of the target population for whom adequate records exist and who are accessible to the researcher. Descriptions of these populations should include whether cases were drawn from an available list or a list GAO developed.

3.   A Description of Limitations

The description of limitations includes those related to nongeneralizable samples in general and to the particular sampling strategy, criteria, or case lists used to select cases, and the reliability and validity of the evidence developed from the review. The following note on the primary limitation of a nongeneralizable sample should be included: "Results from nongeneralizable samples cannot be used to make inferences about a population."

4.   Disclosure That a Requester Suggested One or More Cases or Locations for Review

This disclosure should be included if it is applicable, along with a note as to whether the requested cases fit within the chosen sampling strategy and criteria. (See ARM guidance, [Handling Requester's Suggestions for Locations or Items to Test](#)). Appendix II has examples of report language describing case selection.

NONGENERALIZABLE SAMPLING STRATEGIES

This appendix describes various approaches teams might use in selecting cases. In general, nongeneralizable sampling strategies are purposeful sampling strategies rather than probability sampling strategies—that is, the sampling strategy seeks to select cases in which GAO can gain deeper insight to answer researchable questions. Background work should be conducted to identify the cases, or the types of cases, that are most useful to the engagement.

More than one purpose or criterion may be used in making selections, and in some situations, sampling strategies can be combined. No one approach is better than another. The best choice is determined by the requirements of the research objective. Note that a limitation common to all strategies is that they will not allow you to generalize findings to the larger population. Table 1 lists nonprobability sampling techniques for selecting cases for a purposeful sample.

Table 1: Nonprobability Sampling Techniques for Selecting Cases for a Nongeneralizable (or Purposeful) Sample

Strategy	Description	Example	Some strengths	Some limitations
Purposeful Sampling	A relatively small number of cases is selected to be **illustrative** of program operations under a variety of conditions.	1. Three cases might be selected to ensure some variation in size of facilities, U.S. regions, incidence of reported problems, and old versus new operating procedures. 2. To study passport and visa inspections at U.S. air, land, and sea ports, we might select ports that vary in the number of border entries and the number of fraudulent documents detected.	Can help in interpreting other data; can provide anecdotes and illustrations about program operations under a variety of conditions. Many permutations of cases could provide some variety in the conditions under which programs operate. Does not require a complete population list.	Data collected are anecdotal and while we report the results we find- firmer conclusions could only be drawn through the use of more rigorous data collection/sampling methods. Cannot provide many insights into the effects of any one set of conditions.

Strategy	Description	Example	Some strengths	Some limitations
Stratified purposeful sampling	This is a specific type of purposeful sampling. Cases are selected from within major subgroups, or strata, of the population, to capture major variations, although commonalities may emerge when data are analyzed.	1. In a study of differences in management structures of local housing authorities of various sizes, you might select a few small, a few medium, and a few large authorities. 2. In a study of Federal Emergency Management Agency grants for flood zone management, you might select local governments in a variety of geographic areas, of various sizes, and with a range of problems with flooding.	Allows you to make qualified comparisons between different subgroups of a population and to discuss issues each subgroup faces. You need not have a complete population list.	Increasing the number of criteria, or strata, you want to consider can quickly increase the number of cases you need to sample.
Intense case sampling, or heterogeneity or maximum variation sampling	As in stratified purposeful sampling, cases are chosen that have the greatest variation on key factors in order to describe central themes that emerge across cases with great variation.[a] Developing a matrix of cases and their characteristics can be useful for identifying how they differ and selecting which to include in your job.	For a report on the effects of major wildfires, you might identify fires that burned numerous ecosystems, affected a wide variety of natural resources, and involved multiple federal and state agencies.	Heterogeneity in small samples can be a difficulty for other sampling approaches. Maximum variation sampling overcomes this limitation, since themes emerging across cases capture the core experiences of a phenomenon.[a] Allows you to a describe the context of the issues and interactions of multiple factors. May be less resource intensive; it is like a one-stop-shop for information. Does not require a complete population list	Does not allow you to know whether factors individually have the same effect as they do in combination.

Strategy	Description	Example	Some strengths	Some limitations
Convenience sampling	Cases are selected that are most easily and quickly accessed.	In studying the impact of wildfires on water quality, you might interview attendees at a national conference of hydrologists rather than contacting experts one by one at their offices.	Often requires fewer resources since little preparation is required. Does not require a complete population list.	The sample represents only one segment of the population; the bias this introduces cannot be determined. May not be the most useful strategy to answer the researchable questions.
Critical instance sampling (aka unique)	Cases are selected because they are unique or rare in a population (Three Mile Island, Hurricane Katrina) and of great interest for illustrating a success or problem that probably affects all sites or calls into question a generally accepted assertion. The question might be: What is happening and why? A type of criterion sampling—sampling all cases that meet a predetermined criterion of importance—such as sampling all cases that exceeded an expected timeliness standard to determine why they exceeded it. Can be useful for reviewing program management or management information systems.[a]	In studying major failures of nuclear reactors, you would have few choices but to study Three Mile Island. To see whether federal policies caused problems in port operations, GAO examined the Port of New York, which is diverse and has a high work volume. Problems would be likely to show up at this site and at others; if no problems were observed at the Port of New York, problems were unlikely at other sites.	May be useful if resources are constrained and only one or a few cases can be examined. May allow you to conclude that a problem, challenge, success, or understanding that occurred in cases you collected data from is highly likely to occur elsewhere.	The biggest pitfall in is insufficient specification of the client's question— that is, the approach will not allow you to meet your client's needs if the researchable question seeks to understand the phenomenon beyond a particular case.
Expert referral	Cases are selected by asking one or more experts to list cases that meet the criteria for the population being studied. It is important to document that the "experts" truly are experts in their professional capacity.	In studying the use of federal transportation funds in state mass transit projects, you might ask state and national transportation policy experts to recommend specific projects.	Helpful when you are unfamiliar with a topic and can save time in identifying cases that are appropriate to the subject area. Does not require a complete population list.	The resulting group of cases is subject to the experts' biases, which are practically impossible to define. The approach is only as objective as your instructions to the experts and the experts' awareness of their own biases.
Best or Worst Case Sampling (aka Extreme Case)	Cases are selected that are information rich, because they are unusual or special in representing extremes, outliers, or atypical cases—the best (largest, most expensive, most	Best case approach: In studying the effectiveness of community programs to reduce handgun violence, you might select the five programs that led to the greatest	The best case approach allows you to make statements about possible success, giving an upper bound to the issue or providing a best-	An engagement can include both best and worst cases but their use will not allow you to make statements about typical cases or the ranges of cases. If

Strategy	Description	Example	Some strengths	Some limitations
	efficient, oldest) or worst (smallest, least expensive, least efficient, youngest) instances of a particular phenomenon. An engagement may include both best and worst cases, which is referred to as bracketing. A similar type of sampling approach is Intensity sampling; it follows the same logic but instead of focusing on extreme cases, it selects those that manifest the characteristic intensely but are not so unusual as to distort the phenomenon of interest.[a] Background work should explore the nature of the cases and identify those that well represent the extremes.	drop in firearm fatalities. "In Search of Excellence" identified companies that a group of industry observers considered innovative and excellent.[a] Worst case approach: In investigating errors in Medicare claims processing, you might select medical institutions with particularly high rates of fraud. "When Battered Women Kill" examined the most extreme cases of domestic violence to illustrate this issue.[a] Often forensic audits are designed to detect and identify fraud. In this regard, an appropriate certified fraud examiner (CFE) approach is to select cases in a manner that will maximize fraud identification. Without proper context, this approach to selecting cases may appear to be "cherry picking" the worst examples; however, it may, in fact, be necessary to target the selections in this way to facilitate investigation, illuminate control limitations, and eliminate illegal activity. Because of potential issues about balance/bias, we should provide appropriate overall context for any fraud findings and assessments to ensure that the reader does not mistakenly conclude that the selected cases are more prevalent than they are.	case scenario. It also increases the likelihood that you will notice potential outcomes. The worst case approach allows you to make statements about possible problems, giving a lower bound to the issue or providing a worst-case scenario. It will also allow you to make statements about why a case is not successful.	you select only best cases, you will not be allowed to say anything about worst cases; if only worst cases, you will not be allowed to say anything about best cases. Extreme cases may be discredited as too unusual to produce useful data (intensity sampling or another Approach may work better).[a] When a sample has been drawn to maximize the chances of finding something such as fraud, we need to be clear that the extent of the instances in the sample do not reflect the extent of those in the larger, unsampled, population

Strategy	Description	Example	Some strengths	Some limitations
Snowball sampling	Used when the unit of analysis is a person. Beginning with an initial list of cases, ask each person to refer you to additional persons. The group of referred cases (or "snowball") grows larger and then narrows as a group of individuals are identified frequently. This group becomes the cases from which you will collect data.	In studying access to public services for homeless families, you might identify an initial group of families at a homeless shelter and then ask them to refer you to other families.	Might be the only way to obtain information about a population that is difficult to track down. Since all group members are related in some way, you can study their relationships and interactions. Does not require a complete population list.	It is practically impossible to determine the portion of the population represented by your sample.
Typical case sampling	Cases are chosen that represent the typical instance of a particular phenomenon. Can help profile a program or policy. You can identify cases from many sources—agency staff, key stakeholders, survey or statistical data (using frequency distributions). It is important to ensure buy-in on what defines "typical."[a]	In studying the implementation of welfare reform, you might select states with close to the median per capita amount on welfare outlays.	Allows you to describe issues facing the typical case chosen or about the most likely situation.	Does not allow you to say anything about the best or worst case or the range of cases.

Source: GAO.

[a]Michael Quinn Patton, *Qualitative Research and Evaluation*, 3rd ed. (Thousand Oaks, Calif.: Sage Publications, 2002), pp. 230–42.

Table 2 shows two probability sampling strategies that can be used to select cases for a nongeneralizable sample. Although these are probability sampling strategies, when only using them to draw small samples, the sample size usually does not allow you to generalize to the larger population.

Table 2: Probability Sampling Techniques for Selecting Cases for a Nongeneralizable Sample

Strategy	Description	Example	Some strengths	Some limitations
Simple random sampling	Every case in the population has an equal chance of being selected.	In assessing the internal controls of an agency's inventory system, you might randomly select storage facilities to visit to perform completeness and accuracy tests of on-hand items to the database inventory.	Allows you to select cases while ensuring there is no selection bias. Useful if you have no characteristics or basis on which to choose another approach and have no time to screen cases to identify others better suited to your job.	Does not ensure that specific types of cases are selected and, thus, does not allow you to say anything about cases with particular characteristics. Requires you to have a list of the population from which to select cases.
Systematic sampling	Cases are chosen according to a predetermined strategy (e.g., every X case), which could include stratification.	1. To determine whether agency grant award files contain required documentation, you select every X^{th} file (as listed in the grant award database) for review. With stratification, you select every X^{th} file from each regional office. 2. To gather opinions of visitors to the National Mall, you ask every 5th passerby to complete an interview. With stratification, this could include, for example, every 5th man and woman passing by.	Ensures selection from the range of possible cases throughout the population. Can be used when you have no list of cases from the population.	May result in a biased sample if systematic patterns correspond to your selection strategy.

Source: GAO.

OSM REPORT LANGUAGE DESCRIBING CASE SELECTIONS

After considering the steps described in this guidance, staff should describe in their report how case selection was made. This appendix gives five examples.

1.   We chose the four locations—Louisville, Kentucky; seven counties in New Jersey; Memphis, Tennessee; and Philadelphia, Pennsylvania—because all four appeared to be among the best implemented, most consistently applied, most mature programs in the country. They also offered geographic diversity and were willing to be part of the study, which involved altering their internal processes and procedures somewhat to accommodate the design. By including four locations that were among the best implemented, the evaluation was poised to determine whether family preservation services can be more effective than "regular" services when they are well implemented. In other words, the chances of seeing program success was deliberately increased.

2.   To gain a balance of views from states, we selected a nongeneralizable sample of 10 states—the 5 states that had the most sites proposed to the NPL in the past 5 years (California, Florida, New Jersey, New York, and Texas) and the 5 states that had no sites proposed in the past 10 years (Arizona, Delaware, Nevada, North Dakota, and Wyoming). This sample does not represent the views of the states that did not fall into either group.

3.   To learn more about the Department of State's public affairs operations, we visited U.S. embassies in Cairo, Guatemala City, and London. This ensured that we visited posts that had relatively large, medium, and small public affairs staffs and covered several major regions of the world. While the sample allowed us to learn about many important aspects of, and variations in, the department's public affairs operations, it was designed to provide anecdotal information, not findings that would be representative of all the department's more than 200 posts worldwide.

4.   To evaluate the extent to which policy guidance was applied at selected sites, we analyzed the permit records and other documentation of six selected park units that we visited, and we interviewed Park Service headquarters, regional, and park unit officials. We selected these park units because, during fiscal year 2003, they had issued the greatest number of special event and filming and still photography permits in the six Park Service regions within the continental United States. Because we used a nongeneralizable sample to select the units that had issued the greatest number of permits in fiscal year 2003, our findings cannot be used to make inferences about other park service units. However, we determined that the selection of these sites was appropriate for our design and objectives and that the selection would generate valid and reliable evidence to support our work.

5.   We used a purposeful stratified sampling procedure in which we intentionally chose to interview people with particular characteristics to capture both common core experiences and important variations among those with differing characteristics. We identified the

states in which victims resided before the hurricane (Alabama, Louisiana, Mississippi, and Texas) and whether they received HUD housing assistance before the hurricane as two characteristics that would influence victims' needs and their experience finding housing. When using a nonprobability sample, like a purposeful stratified sample, it is important to be resource efficient in data collection but also to simultaneously collect enough data to ensure saturation, or repetition, in the information obtained.[7] Therefore, we initially planned to collect data from 48 victims—24 who had received public housing assistance before the storm and 24 who had not—and, within both groups of 24, from an equal number of participants from the four states (see table 1).

Table 1: Number of Completed Interviews with Hurricane Katrina Victims

State	HUD housing assistance before the disaster	
	Number who received	Number who did not receive
Alabama	2	6
Louisiana	8	6
Mississippi	8	7
Texas	0	1

Source: GAO.

Note: We attempted to complete 6 interviews per table cell.

We identified victims by word of mouth and from HUD and FEMA disaster assistance lists. Names and contact information for victims identified by word of mouth were provided to us by organizations working directly with victims, such as churches and nonprofit organizations, and by other victims. A list that HUD provided us as of July 14, 2006, was our primary means of identifying victims who received HUD housing assistance before the hurricane. FEMA provided us a list as of July 20, 2006. After eliminating cases that had no telephone numbers, we systematically selected victims' names from the agencies' lists.

We contacted victims and asked them to participate in our telephone interview, which lasted approximately 60 to 90 minutes. If a victim could not be reached, declined, or was not available at the scheduled interview time, we eliminated the name from our contact list. We completed 38 interviews with disaster victims. We contacted approximately 323 victims to request their participation. Demographic information on the victims we interviewed appears in table 2.

---

[7] Janice M. Morse, "Designing Funded Qualitative Research," in Norman K. Denzin and Yvonna S. Lincoln, eds., *Handbook of Qualitative Research* (Thousand Oaks, Calif.: Sage Publications, 1994), pp. 220–34; Patton, *Qualitative Research and Evaluation Methods*, 3rd ed., pp. 245–46.

Table 2: Self-Reported Demographic Data on Disaster Victims Interviewed

Demographic element	Years	Number
Age		
Average	49	
Range	22–72	
Refused		1
Ethnicity		
Black or African American		26
White		9
American Indian or Alaska Native		2
Hispanic or Latino		1
Asian		0
Native Hawaiian or Other Pacific Islander		0
Homeownership		
Renter		25
Owner		13

Source: GAO.

Results from nonprobability samples cannot be used to make inferences about a population, because in a nonprobability sample, some elements of the population have no chance or an unknown chance of being selected as part of the sample. Our findings cannot be generalized to all victims of Hurricane Katrina, but when coupled with results of our group of experts, interviews with agency officials, and housing advocates, they do provide useful insight into the experiences and needs of victims of this disaster.

SAMPLE DOCUMENTATION WORKPAPER

Prepared by: Type name here
Date Prepared: Type date here
Reviewed by: Type reviewer name here
Job Code: Type job code here

Index: Type bundle index here
DOC Number: Type document number here
DOC Library: Type library name here

Record of Nongeneralizable Sample Decisions

Title1   **Summary of nongeneralizable sample selection decisions for (name of job)**

Purpose: **Document decisions and steps taken to select a nongeneralizable sample.**

Purpose of the nongeneralizable sample:

Design matrix questions the sample will help answer:
(Describe how the sample will help answer researchable questions in your design matrix. Explain why a nongeneralizable sample is appropriate—e.g., to describe aspects of an issue, understand the context of a problem, provide anecdotes about a problem. Consider explaining why a statistical or generalizable sample was not appropriate.)

Data sources:
(Describe the data sources or lists from which the cases were selected. If it is appropriate, discuss data reliability issues. If appropriate, attach with the data sources or lists a spreadsheet that clearly indicates which cases were selected and which were not.)

Sampling strategy:
(Describe the strategy used—e.g., a convenience or intense case strategy. Appendix I of ARM's guidance on selecting nongeneralizable samples lists selection strategies to consider.)

Cases selected and criteria for selection:
Approximate number of cases in population ____
Number of cases selected _____
(Provide the rationale for the number of cases selected.)

List cases selected:
(Provide the criteria for the cases selected and reasons why other plausible candidates were rejected.)

Strengths and limitations of the case selection strategy:
(A limitation common to all strategies is that they cannot be used to generalize findings to a larger population. Appendix I of ARM's guidance on selecting nongeneralizable samples gives other examples of strengths and weaknesses.)