
COLLECTING DATA IN THE FIELD

Demetra Smith Nightingale, Shelli Balter Rossman

There is much interest among policymakers, public officials, the media, and the general public in understanding what actually occurs in programs at the ground level and learning what kinds of services or programs seem to work best for different target groups or in different localities. Evaluators are often called on to examine how local programs or agencies operate and how services are delivered. The inquiry might be part of a multifaceted formal evaluation, or it might occur independently as a self-contained study. The focus of analysis might be one program in a single location or several programs in multiple locations, and the study might be cross-sectional (conducted at one point in time) or longitudinal (addressing operations over a period of time). In addition to formal evaluations, federal and state officials and program managers routinely visit local programs to get a better sense of operational realities. Public officials with monitoring responsibilities, for example, visit programs to review specific issues.

Many public and non-profit program evaluations are field based to some extent, meaning that researchers generally collect some information at locations where programs are operating. The field location may be a federal, state, or local agency office, a nonprofit organization, or the office or facility of a public or private operator or provider of services. While on site, evaluators systematically collect information by observing program activities or through surveys; focus groups with participants; interviews with officials, staff, or other stakeholders; and retrieving data from management systems or case file reviews. There are many strategies and various methods for analyzing the qualitative

and quantitative data collected. Among the more common types of studies or activities, for example, are single-site case studies, organizational studies, and thematic cross-site comparative studies. Field-based data collection is also frequently used in process studies, implementation analyses, or organizational assessments, as well as outcome and net-impact evaluations.

Other chapters in this volume describe specific data collection methods such as questionnaires, focus groups, and interviews; designs such as case studies, implementation analysis, and logic models; and quantitative and qualitative analytical methods and tools that typically use data collected in the field. This chapter covers some of these topics and complements other chapters by focusing directly on the rationales for collecting data in the field, the different models and conceptual frameworks that can guide field studies, and the procedures and logistics associated with collecting and maintaining data and ensuring quality control. Examples presented include multisite studies that are part of formal evaluations, as well as separate process studies and less extensive efforts that are more appropriate for routine program monitoring and oversight.

Objectives of Field Studies

The details of fieldwork depend on the objectives of the data collection, which are based on the overall purposes of the study or project within which the fieldwork occurs. It is important to fully understand what the fieldwork is intended to achieve, how it fits into the conceptual framework of the evaluation as a whole, and the categories of information it is expected to collect.

The objectives of the fieldwork determine both the focus (priorities) and the scope (intensity) of the data collection activity. At least two types of fieldwork studies are common: program management projects and program evaluation. Box 17.1 displays some examples of objectives from field-based studies that address issues that are related to both management (e.g., program performance, service delivery assessment) and evaluation (e.g., implementation and process analysis) objectives.

Box 17.1. Examples of Field Research Study Objectives

Objectives for a Multisite Adult Drug Court Evaluation (Rossman and Roman, 2004)

National web-based survey objective	Document extant adult drug court models as context for selecting representative sites to participate in impact evaluation.
-------------------------------------	--

Process evaluation objective	Document and assess drug court implementation practices—including courtroom observation, procedural manuals and participant records reviews, and staff and stakeholder interviews and focus groups—in twenty-three adult drug court sites.
Impact evaluation objective	Compare relapse and recidivism and also psychosocial outcomes of drug court participants to outcomes of a comparison group of offenders with similar drug and crime histories in six non-drug court jurisdictions.
Cost-benefit analysis objective	Determine drug court costs and benefits as compared to “business as usual” in criminal justice responses to substance abuse offenders.

Objectives for an Implementation Evaluation of Welfare-to-Work Programs in New York City (Nightingale, Pindus, Kramer, Trutko, Mikelson, and Egner, 2002)

Performance analysis objective	Examine program participation and outcomes over eight years.
Implementation analysis objectives	Describe the organization, management, and service delivery procedures in local welfare offices and in programs under contract to serve welfare recipients. Identify policy, bureaucratic, and political factors that influence the way local programs are structured and managed.

Objectives for a Program Outcome Evaluation of a Child Support Enforcement Collections System (Holcomb and Nightingale, 1989)

Performance analysis objective	Estimate the impact of the information clearinghouse in five demonstration states on child support collections and government savings.
Implementation analysis objectives	Document and assess how the clearinghouse was planned, implemented, and operated. Assess the feasibility of implementing and operating an automated clearinghouse by identifying problems encountered and solutions applied. Identify differences in the clearinghouse across the five demonstration states and reasons for the differences.

(Continued)

Objectives for an Impact Evaluation of an Aftercare Program for Substance Abusers (Rossman, Roman, Buck, and Morley, 1999)

Impact analysis objective	Compare substance abuse relapse and criminal recidivism of offenders randomly assigned to receive program services (Opportunity to Succeed [OPTS] clients) to the same outcomes for offenders not enrolled in the program.
Implementation analysis objectives	Document and assess the implementation of collaboration between probation and parole entities and service providers.
	Document and assess the nature and extent of case management and core services (substance abuse treatment, employment, housing, medical and mental health treatment, and family support) provided to OPTS clients.

Program Management Fieldwork Model

Federal, state, and local program managers routinely conduct monitoring reviews that involve field visits, reviews of records, interviews, and observations. The topics or issues reviewed depend on the needs of management (such as determining compliance with regulations or improving program performance). Analysis may be quantitative or qualitative, ideally based on predetermined management standards or criteria. The fieldwork is usually conducted by managers or staff of public agencies. However, in some cases, contractors may be engaged to carry out the management review: for example, as part of a performance monitoring project or an assessment of technical assistance needs. The review may involve obtaining management data from computerized or hard copy case files or data in a management information reporting system. The results, typically presented in site reports, may lead to recommendations for corrective action or performance improvement.

Program Evaluation Fieldwork Model

Evaluators ideally collect information on predetermined topics. The classification of topics is based on the overall evaluation project and its objectives. Various data collection methods might be used, including interviews, surveys, focus groups, observations, statistical compilations, and record reviews. Standard social science principles (such as validity, reliability, and objectivity) must be considered in developing the fieldwork plan. The fieldwork and the evaluation are based on theoretical models and hypotheses. Both qualitative and

quantitative analysis may be conducted. Individuals who have academic or professional training in research or evaluation usually conduct the fieldwork. Evaluators can be staff of public agencies or researchers from outside research organizations or universities. The results of the work are presented in project reports and often are integrated into other components of a larger or multi-component evaluation.

Each of these and other fieldwork models—or reasons for conducting the study—can potentially involve similar types of data collection methods, but each is based on somewhat different professional practices and experience. The important point is that the specific objectives of the fieldwork set boundaries or standards for the data collection effort. Although it is not essential that a study have a clearly defined fieldwork model, one usually exists, even if it is unstated. The fieldwork model heavily influences specific details about how the fieldwork is designed, the types of data collection instruments used, the professional backgrounds of data collectors and analysts, and the types of quality control and analytical methods employed. These issues are discussed in the sections that follow.

Thus, fieldwork is conducted for at least two purposes:

- To describe what happens at the level being examined (local office, local program, local agency, local community, state office, state agency, and so on) by collecting information about procedures and data on activities, services, institutional features, outcomes.
- To explain why the situations are as they are.

The specific objectives of a fieldwork effort usually fall under one or both of these two general purposes. Before researchers design the details of the fieldwork, it is critical that they articulate clearly and specifically the evaluation questions and issues that relate to the fieldwork portion of the study. Some field studies have very specific objectives, even though the overall evaluation may address broader issues. In contrast, some field-based components of evaluations are called on to address broader program issues, while other components focus on specific questions.

For example, in most large-scale program evaluations that estimate client impacts at the individual level (such as the effect a program has on individuals' employment or educational achievement), an implementation analysis component of an evaluation may involve field data collection to document and investigate in detail specific characteristics of the program being evaluated, such as organizational structure, intake procedures, management functions, and staff job satisfaction. A complementary impact analysis component the

evaluation would focus on statistically estimating the change in individual outcomes. Thus, many program participant impact evaluations commonly include process analysis or implementation analysis components with fieldwork to document specific details of a program. The qualitative program descriptors collected in the field can then be transformed into quantitative program variables and incorporated into statistical analyses of program impacts on individual clients to explain the impact findings more fully.

In most evaluations, it is necessary to build the fieldwork design around the basic evaluation questions. Some evaluations attempt to understand variations in performance. An early institutional analysis of the Work Incentive Program (WIN) for welfare recipients, for example, was designed to examine the organizational, managerial, and service delivery characteristics of high- and low-performing state and local programs to determine what seemed to be related to differences in performance (Mitchell, Chadwin, and Nightingale, 1979). The resulting information was used to develop performance improvement strategies for the program nationally. The study had two components: a quantitative analysis of program performance and a more qualitative analysis of features of high- and low-performing programs. The second component relied heavily on information obtained by teams of evaluators who conducted fieldwork in forty-three local communities. Other field-based studies are designed to document variation across sites or identify potentially promising approaches. A study of the implementation of expanded summer jobs programs for youth conducted fieldwork in twenty sites to identify potentially promising approaches that could be of interest to other localities (Bellotti, Rosenberg, Sattar, Esposito, and Ziegler, 2010). These types of issues determine several aspects of the design of the fieldwork.

Design Issues

When the evaluation questions, objectives, and issues of interest have been clearly specified, the evaluators need to make a number of design decisions:

- Determine an appropriate method for guiding the data collection and subsequent analysis.
- Select sites.
- Decide which types of data collection instruments to use and then develop them.
- Identify and select respondents.

These decisions depend greatly on any cost and time constraints that may exist. Evaluations addressing similar or even identical questions may use different fieldwork designs reflecting different cost and time constraints.

Frameworks for Guiding Data Collection

It is sometimes tempting to attempt to examine all aspects of an agency or program using unstructured data collection methods—often described as “getting into the field and finding out what is going on.” Even though one might obtain a valuable sense of “what is going on” from a study like this, it is not based on a conceptual framework, it lacks methodological rigor, and the credibility of findings reported may be compromised.

However, one of the greatest pitfalls in conducting field-based studies is the risk associated with collecting too much information; not only can this consume resources unnecessarily, but also analysts can easily become overwhelmed with mounds of qualitative and quantitative information: field notes, interview transcripts, focus group reports, management reports, and site reports, among other data. Unless the data collection stage is well organized, the analysis will be very difficult and subject to problems of accuracy and reliability. To avoid subsequent analytical problems, it is important to use or develop guidelines or a framework at the beginning to help focus the study.

Just as there is no common set of research questions that field-based studies are called on to address, neither is there a common framework used to guide the data collection—in other words, there is no cookie-cutter framework that can be adopted. Instead, the evaluators must develop an overriding framework for each study. In some evaluations, the guidance can be based on the research questions, using them to structure the data collection. Many studies use graphic logic models to help structure data collection. In large-scale evaluations, more theoretical conceptual models often are used.

Research Questions. Here is a sampling of the many types of research questions that might be addressed in studies that are likely to involve fieldwork:

- What are the major goals and assumptions underlying the policy that was adopted? What are the policy's underlying premises and assumptions? What is the policy intended to accomplish? How does this vary by level of program (for example, state, local)?

- What are the main program outcomes and performance of a program or policy? How are outcomes and performance measured? What are the priorities among measures? How consistent are the various outcome and performance criteria? What is the trend in performance over time or across sites?
- What are the organizational and service delivery structure and context in which the policy is operationalized? How is the organization structured? What are staff roles and responsibilities? What organizational arrangements and linkages are in place to deliver services? What types of interagency and interprogram interactions and collaborations are involved?
- How are key management functions carried out, and what role do they play in the program? How is program planning structured? Who is involved? What types of management information are used and for what purposes (planning, monitoring, performance analysis, performance improvement, evaluation, or something else)?
- Is the program following the formally established strategy? Are all the components implemented as required? Are all the components implemented efficiently? If linkages among components are necessary, are they all in place? Are some components weaker than others?
- Do programs deviate from the planned model? If so, what are the reasons for adaptations/deviations?
- How are services delivered, and how do clients flow through the service delivery system?

Research questions such as these are commonly posed to evaluators. A small-scale field-based evaluation might address one specific type of issue—for example, how key management functions, such as program planning, performance measurement, or staff development, are carried out in a particular program. The important dimensions of that general issue can be clarified in discussions with policymakers, program administrators, or agency officials before finalizing the actual data collection instruments. Then the evaluators can specify the types of questions or data items that will have to be collected and the types of respondents in the field who might be interviewed. Thus even small-scale studies focusing on one or just a few localities and on a few related issues should develop a data collection and fieldwork plan by carefully specifying the various dimensions of the evaluation questions at hand. The key point is that it is important to have a clear guide for collecting information in the field, even if the evaluation question seems simple and straightforward.

A study plan can help the evaluators maintain objectivity and avoid over-collection of information. A plan based on research questions should, at a minimum:

- Clarify each of the evaluation questions to be addressed in the study.
- Identify types of information required to address each question (for example, program procedure information, program data on outcomes, organizational information, staff perceptions on key issues, customer satisfaction).
- Specify data collection strategies to use to collect each item of information (for example, management information data, staff surveys, administrator interviews, and customer or user surveys).

Logic Models. At a somewhat higher level of methodological sophistication, such as an evaluation that involves fieldwork to multiple sites, evaluators often benefit from preparing a diagrammatic or graphic model to specify the key dimensions of the issue being addressed and how different dimensions or factors relate to each other and to the research questions and evaluation outcomes. Flowcharts and logic models have long been used by public administrators to plan and develop programs, specifying program components, client-flow procedures, management activities, and program outcomes. (See Chapter Three for more detail on logic models and their use in evaluations.) Logic models are used for developing programs and delivery systems to improve the quality of services, such as to ensure that mental health service treatments or interventions are appropriate and consistent with clinical practice (Hernandez, 2000). Decision points and action sequences are included in models, and some models indicate when different levels of a program or organization interact around a particular activity or service. Carefully developed logic models have the potential to improve service provision or program management because the model, or plan, incorporates a theoretical understanding of how different actions or steps interact to produce certain outcomes.

Just as flow models have been routinely used in program planning development and administration, they are now increasingly used for program evaluations of management issues, as a complement to individual impact evaluations, and for addressing other issues that involve sequential phenomena (Abbott, 1995). Logic models can help focus the topics that should be included in field interviews, for example, documenting the organizational units, the sequencing of activities, and outcome or performance measures. In formative evaluations, if outcomes are less than acceptable, staff and managers can use the logic model to diagnose problem points and suggest improvement strategies.

Implementation Models. Large-scale evaluations of public programs generally include implementation analysis components in addition to individual impact analysis, cost-benefit analysis, and program outcome analysis. Program implementation components of large evaluations often focus on the details of program processes: understanding the internal dynamics and structure of a program, the organizational context in which the program operates, the ways clients enter and move through the program, and the ways the program is structured and managed. Describing and analyzing the process involves delineating program services or client activities into discrete components, documenting how these components fit together in terms of client flow, and obtaining a variety of perspectives from people inside and outside the program on the strengths and weaknesses of the various components. Process analysis is considered a subcategory of implementation analysis, focusing on the specific procedures (such as provision of services and client flow) that occur at the operational service delivery level.

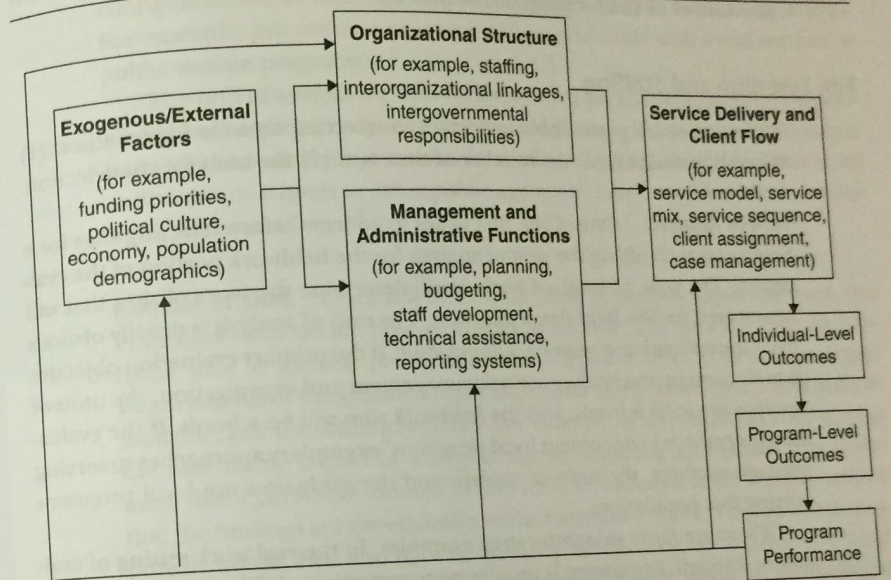
Using graphic depictions similar to logic models, program implementation and process studies that are embedded into comprehensive evaluations often use conceptual framework models to guide the development of hypotheses, data collection, and analysis. Implementation analysis draws from many academic disciplines, especially those related to organizational behavior, social networks, economic behavior, and group dynamics.

Implementation studies of many human services programs, such as welfare-to-work programs, employment services, social services, and programs for formerly incarcerated individuals, draw extensively from organizational theory and systems theory to build a conceptual framework that defines the various categories of information that will be collected in the field. Figure 17.1 shows a simplified depiction of an implementation framework used in impact evaluations, documenting and assessing factors that influence program outcomes (Rossman, Roman, Buck, and Morley, 1999).

The general premise is that some factors (such as the economy, funding levels, or political priorities) totally outside the control of managers and administrators affect how a program is structured and designed. The decisions about structure, design, and operations in turn influence the nature of service delivery and, ultimately, program outcomes.

Evaluating the implementation of a program or policy therefore requires documenting each factor or component that is hypothesized to influence outcomes. A carefully specified framework that defines the factors in each category can form the basis for organizing data collection instruments and analysis to explain how a program operates.

FIGURE 17.1. SIMPLIFIED COMPONENTS OF AN IMPLEMENTATION ANALYSIS CONCEPTUAL FRAMEWORK.



Source: Rossman, Roman, Buck, and Morley, 1999.

Large implementation studies based on theoretical understanding of how organizations and programs operate stand in contrast to less formal efforts that do not use formal fieldwork protocols and involve visiting a few convenient sites, meeting with local administrators, touring a program, and possibly speaking with clients. Some theoretical models from certain disciplines such as economics or sociology can be adapted to serve as conceptual frameworks for field studies. Systems theory and program logic models can also be used to establish conceptual frameworks for field-based studies.

Program implementation is complex. To fully understand what is happening in a program, the evaluator must examine it from different perspectives and view each component of the program both separately and as part of the whole entity. Without a guiding framework, it is very easy to become overwhelmed in the details and lose track of the program as a whole. Conceptual models and frameworks help the evaluator stay organized and avoid information overload. There is no one standard conceptual framework. Each

study requires developing a framework that draws on relevant intellectual theory from established academic disciplines or on accumulated knowledge from past studies of similar programs or policies.

Site Selection and Staffing

Three issues are particularly important in selecting sites for field studies: (1) the unit of analysis, (2) the number of sites, and (3) the basis for site selection.

Unit of Analysis. One of the first issues to address before selecting sites for a field study is clarifying the unit of analysis for the fieldwork portion of the evaluation. The unit, or level, of analysis will determine the types of sites that will be selected for the field data collection. The unit of analysis is usually obvious from the evaluation questions. For example, if the primary evaluation objective is to document and analyze school management and organization, the units of analysis are local schools, and the fieldwork sites will be schools. If the evaluation objective is to document local programs' exemplary approaches to serving teenage mothers, the units of analysis and the study sites are local programs serving this population.

These are fairly straightforward examples. In the real work setting of evaluation, though, the choice is usually more complicated. There are often multiple dimensions to evaluations that require different levels of analysis. For instance, if the evaluation objective is to determine how teenage mothers feel about their circumstances and the services available to them, these mothers are the units of analysis. But the evaluators must decide how—that is, from what source—to identify the mothers: schools, hospitals, local programs, welfare rolls, or cities or states. These then also become units of analysis. The final analysis might focus on mothers served by local programs in general or on each local program (with the mothers each of them serve) or on both levels of analysis. Thus, if evaluators want to be able to discuss individual programs in their analysis, they should consider this when selecting the sites where the fieldwork is to be conducted—for example, a city, a neighborhood, or one or more institutions or programs serving the city or neighborhood.

Multiple levels of analysis are common in program organizational evaluations, that is, evaluations not focusing just on individuals. In a national evaluation of coordination between welfare programs and job training programs, for example, a number of units of analysis are possible: local communities (within which all job training and welfare programs would be examined), states (within which all job training and welfare programs would be examined), one or more specific job training programs (which could be examined at the state or local

level), or one or more specific welfare programs (which could be examined at the state or local level). Evaluators often also look at different partnership configurations and interorganizational arrangements as the unit of analysis. For example, job training programs often collaborate with social services or public welfare programs).

The unit of analysis for fieldwork studies should be obvious given the overall evaluation questions specified in the study design. Common units of analysis in evaluations are local programs, local offices, individual local facilities (such as libraries or schools or other public agencies), cities, neighborhoods, institutions, and states.

Number of Sites. Once it is clear what units of analysis should be used, the evaluators must decide how many sites to include. The final decision on how many sites to include in the field study depends heavily on the analytical requirements of the specific impact evaluation being conducted, the resources available, and the staffing required. For example, if an evaluation is estimating the nationwide impact of a large program, sampling statisticians may determine that a particular number of sites must be included in the study to ensure that the findings are generalizable to the nation as a whole. The fieldwork component would then also have to include a certain number of sites (either all sites in the overall evaluation or a subsample of sites, based on the statistical sampling parameters). In most field studies, decisions about staffing and site selection are made simultaneously.

Four main factors affect the resource levels required for fieldwork:

- Travel distance.
- Length of time on site.
- Level of evaluation staff required.
- Number of evaluation staff required.

The importance of each of these items varies somewhat in different studies. Travel times, for example, may be associated with cost, but sometimes distance alone does not equate to higher costs. Traveling from the east coast to the west coast may cost less and take less time than traveling from either coast to the northern plains. The cost of each of these factors will also be governed by the intensity of data collection. If the field efforts are exploratory, involving unstructured data collection activities, such as discussions with key officials or staff in a program, then each site visit can probably be limited to a short period of time—one or two days—when one evaluator works alone. That

person should be fairly senior to ensure that the exploration is as comprehensive as possible. At the other extreme, an evaluation that involves collecting detailed descriptions of program operations by surveying or interviewing a number of staff in each site will require more days on-site, probably more than one staff person, and a longer period of time.

Most large federally sponsored program evaluations now include an implementation, or process, analysis component. Typically, the fieldwork design involves a team of two evaluators, with one fairly senior and the other either a midlevel evaluator or a research assistant. The two-person team is on-site for three to five days, depending on the size of the site or city and the scope of the inquiry. Two-person teams have proved to be effective for collecting accurate data and, subsequently, analyzing and interpreting the information. The evaluation team can discuss issues and share contextual insights while in the field that greatly strengthen the overall quality of information.

Tight resources may preclude having two-person teams. There are lower-cost staffing configurations that may be fully satisfactory for some less-intensive field-based studies, such as program reviews, survey administration, or collection of routine statistical data. As an example, one public official (such as a state program monitor or program administrator, or a federal inspector or monitor, or an evaluator) can visit a local program for just one day and collect a substantial amount of information. The on-site time can be spent efficiently, using carefully developed data collection instruments, scheduling activities in advance, and following field protocol established before the visit.

Many different activities might be carried out while on site to obtain the information and data specified in the study design based on the objectives and conceptual plan. Site visits can be demanding, and caution should be exercised to avoid placing too high a workload burden on the field staff. To provide a rough estimate of the staff resources required on site, two trained evaluators working together can be expected to accomplish any one of the following types of activities in one day:

- Each person can usually conduct three or four one-hour interviews with staff, administrators, or community officials, allowing time for preparing for each interview and reviewing notes before starting the next interview.
- The team together can conduct two or three (depending on the study) focus groups lasting one to two hours each.
- Each evaluator can administer six forty-five-minute, in-person questionnaires.

- The team can review case records. A typical single welfare case record review, for example, takes between fifteen and forty-five minutes, depending on the size of the file and what information needs to be extracted from the files.
- The team can perform a well-planned combination of two activities, such as one or two in-person interviews and one focus group session, during the same day.

Evaluations generally involve using multiple data collection methods on site, such as trained observer ratings or statistical data collection from management information systems. Thus, the resources devoted to each site visit depend on the mix of activities to be conducted. Fieldwork can become quite expensive. For example, on-site fieldwork that involves collecting data on programs in ten cities, each in a different state, using a two-person team (one senior evaluator and one midlevel evaluator) on site for one week per city, would require about one hundred person-days, at a cost of between \$100,000 to \$150,000 for labor, travel, per diem, and expenses. An additional one hundred to two hundred person-days would be required for previsit preparations and post-visit report writing. Some data collection activities may have additional costs; focus groups, for example, often involve costs for providing individual financial incentives, or reimbursing participants for travel expenses, or arranging for refreshments to motivate participation.

Basis for Site Selection

When there are several possible units of analysis, decisions about site selection are typically based on how the information acquired in the fieldwork will be used by the evaluators. If the purpose is to prepare case studies, each of which can stand on its own, it is not necessary to select sites that are all of the same type. An examination of coordination, for instance, could include one or more local communities and also one or more states. If the purpose is more analytical—perhaps to examine factors that encourage or discourage coordination between two programs—then the evaluators should select sites that represent as broad a range as possible of the various types of programs and situations, perhaps including programs of varying sizes and in both urban and rural locations or programs serving populations with different demographic characteristics.

The sample of sites for fieldwork can be selected in a number of ways. At one extreme, the sampling method may be random, using standard probability

sampling techniques. This requires identifying a universe of possible sites, clustering or stratifying the sites on the basis of some criteria, and then randomly selecting within the strata or clusters. At the other extreme, the site selection process can be purely purposive, with specific sites or specific types of sites chosen, such as small rural sites with high-poverty populations or award-winning programs that are considered exemplary or large programs in high-growth economic labor markets.

Site selection for most field studies usually falls somewhere between these extremes. Evaluators examining exemplary program models might choose sites based on some feature of the program that is of particular interest—the specific populations served, unique locations, innovative program models, and special organizational structure—and choose randomly from among sites meeting those criteria.

In large part, the selection of sites depends on whether the findings from the field are intended to be representative of some larger group of programs or sites or whether they are to be used for stand-alone case studies. If the data and information from the field sites are to be generalized to a larger group of sites or programs, the selected sites should be as representative as possible of the population of sites from which the sample is drawn. If the information is to be used primarily for descriptive case studies and is intended to be illustrative only, then purposive sampling is sufficient.

Even if evaluators choose sites purposively, the selection should still be based on clear guidelines and criteria. In many cases these resemble the types of criteria used to select sites by random stratification or clustering. Examples of selection criteria include level of program performance, rural versus urban location, level of client income, level of client ethnic concentrations, labor market conditions, and geographical location.

In some field studies, site selection might evolve through the evaluators' soliciting interest from local jurisdictions, programs, or agencies. One Urban Institute study was designed to develop and then evaluate a management-oriented performance improvement model in state Work Incentive Programs serving welfare recipients (Nightingale and Ferry, 1982). The study could include only two states, and the following conditions were used to select them:

- The state agency had to have a strong potential for improvement while not currently performing at full capacity.
- The program administrators at the state level had to express a deep commitment to improving their operations.

- State officials had to be willing to participate actively in developing and implementing improvement strategies by making key staff available for the duration of the two-year project.

In the selection process, the evaluators compiled information showing how well each candidate state met these three criteria. The information was obtained through reviews of program performance reports and conversations with key state administrators.

There are no hard-and-fast rules about how to select sites for fieldwork studies. Site selection evolves from the general evaluation objectives. The evaluator must decide whether sites should represent maximum variation or maximum similarity. Regardless of how scientific the site selection process is, evaluators must have pre-established criteria that can also be used subsequently when reporting the implications of the findings. The selection of sites should be based on the objectives of the evaluation, but the final decisions must also consider the staffing that will be required and all costs at each site.

Types and Scope of Instruments

Except in the most exploratory type of fieldwork, evaluators will need to use one or more data collection instruments. At a minimum, the evaluation will need a field data collection guide that includes instructions for obtaining information from interviews, observations, surveys, case reviews, and focus groups. Instruments vary from highly structured to very unstructured.

The structured types of data collection instruments are best known and include surveys, questionnaires, tests, and data coding sheets. Structured instruments have specific items, questions, or codes that data collectors must use in recording information. The least structured evaluations may have no formal data collection instruments. Between are the semi-structured data collection instruments, which consist of topical areas or subject categories, along with questions that the interviewer may use, as well as suggested wording for asking about key issues. (See Chapter Nineteen for a more detailed discussion of semi-structured interviews.)

Field Visit Protocol

Fieldwork projects require careful attention to many procedural and logistical details before, during, and after the site visits. This section discusses the protocol—the critical procedures—that should be developed for a field visit.

Field evaluators should be fully trained on these details. Procedures should be followed precisely to ensure that the information collected is of high quality, the different evaluators collect information in a comparable manner, the fieldwork is minimally intrusive, and confidentiality and other human subjects protections are maintained to the maximum extent possible.

Previsit Preparations

The successful completion of the on-site portion of an evaluation that includes fieldwork depends critically on careful preparation before the site visit. Evaluators should not underestimate the importance of the previsit activities. Previsit preparations include a variety of activities, from setting up site-specific files of existing materials to handling logistical arrangements and recruiting and training field staff.

During the early stages of planning an evaluation, materials should be assembled from a variety of sources, such as government program files, agency and program Internet websites, grant applications, existing databases, or site narratives from prior field trips or evaluation files. Where feasible, these materials should be organized into files associated with each of the designated field sites. A log of contacts (such as phone conversations with the key contact person or program director in each site) can also be included in each site's master file or folder. A log of contacts is especially important in studies that will be conducted over a long period of time and those involving several researchers because it serves as the official record for the study. For evaluations in which many documents are being collected from each site, attach a checklist of materials requested to the site's master folder or file. The materials should be checked off as they are received. Follow-up requests should be made and noted on the log. All files and contacts must be carefully documented to allow the evaluation to be completed efficiently regardless of changes that might occur within the project (for example, if the evaluation team changes due to turnover or if evaluator assignments change).

Site Clearances. Initial contact with field sites should identify any constraints that might affect scheduling or data collection. For example, to gain entry into schools and speak to teachers and staff, evaluators may need to obtain clearances from high levels in an agency, or there may be other evaluations under way in the same site that might require coordinating schedules. Similarly, to conduct interviews in jails or prisons, permissions need to be obtained from facility directors, and should cover such issues as interviewer dress codes,

whether electronic equipment such as smart phones and laptops are permitted, and interview space that permits confidential exchanges outside the line-of-sight of facility staff. If entry clearance or interview authorizations are needed, the evaluator should clarify who is responsible for obtaining approvals (that is, the evaluator or the contact person in the agency), whose permission must be sought, and what information is needed to facilitate the process. Scheduling must be sufficiently flexible to accommodate delays due to bureaucratic obstacles; at the same time, planning should include actions that can be taken to minimize schedule slippage.

Scheduling Visits and Interviews. Several factors, including travel distance and level of staffing required, affect site scheduling. Early communication should identify primary contact persons at selected sites who can serve as liaisons to the evaluation and tentative dates or time periods or possible alternative schedules for the visit. This scheduling will permit advance planning of logistics, such as travel reservations and field staffing assignments. Economies of scale in both travel and staffing costs can often be achieved when visits to geographically close sites can be scheduled together.

Information packages should be assembled and sent to the local contact person to provide background about the evaluation. These should include:

- An overview of the evaluation objectives and of the scope of each field visit.
- Assurances that confidentiality procedures will be followed (for either individuals interviewed or for the site as a whole, depending on the study).
- A sample schedule that the evaluation team would like to follow, identifying those individuals the team wants to interview or meet with, for how long, and the times each day that the field evaluation team members will be available.

Either the field visit team or the designated site contact person can schedule the interviews. The division of responsibility should be clearly established as soon as possible. During that discussion, the evaluators should review with the contact person the list of potential respondents, verifying that appropriate categories of staff have been identified and identifying other persons who may also be important to interview. Usually, the initial interview in a site should be with the key contact person, who can provide an overview of the system.

Evaluators must decide whether they will conduct individual or group interviews and determine the appropriate setting for conducting the interviews. If interview topics are sensitive or there is a need to ensure individuals' confidentiality, one-on-one interviews rather than group interviews should

be planned, with private rooms secured for each session. Even when confidentiality is not an issue, reasonable efforts should be made to secure quiet, unobtrusive settings for interviewing in order to minimize distractions, which can reduce the quality of responses.

Once the field visit has been scheduled, personnel at the site should be notified by phone, mail, or e-mail of the dates for the visit and the time scheduled for the interviews. Several days prior to the visit, contact should be made with the key contact person and possibly with each respondent to reconfirm plans, review the proposed agenda, and ensure the scheduling of the interviews.

Defining Information Needs. Evaluators often collect program materials, agency reports, and other documents related to the study site. It is important to request copies of relevant reports in advance of the actual visit. During the initial telephone conversations regarding the site visit schedule, for example, evaluators should ask about reports, organizational charts, procedure manuals, and other program descriptions, and request copies for review before the visit. If the evaluators intend to collect copies of records or documents while on site, they should discuss their needs in advance with the key contact person or respondents and encourage site personnel to assemble the information before the visit. This is especially helpful when a large number of documents or files are needed or when the site has limited resources, such as a small number of staff or limited access to duplicating equipment, that might make it difficult to compile requested information.

If the evaluation is assessing processes fidelity, or analysing outcomes associated with promising or "best" practices or curriculum-based programming, evaluators should request that training and quality assurance materials, as well as the practice protocols or curricula, available for on-site review, if it is not feasible to obtain copies in advance.

Staffing Assignments. Decisions about the division of labor should be made as early as possible so that field staff know which sites they will cover, which interviews they will be responsible for conducting, and the specific issues they will be exploring. Once field assignments are established, staff should review materials already on hand, such as organizational charts, management information reports, grant applications, program planning documents, and fiscal forms. At the same time, staff should review the research instruments/protocols and the checklist of requested documents, noting which materials still need to be collected, either in advance or when on-site.

The issue of field staff safety bears special mention. Both real and perceived risks of working in some communities (such as high-crime neighborhoods) can make it difficult to hire qualified researchers to staff certain projects. It is crucial to consider the kinds of actions that may be taken to ensure staff safety. Here are some helpful strategies:

- Use two-person teams for site visits in high-risk areas, possibly including someone on the team who is familiar with the local situation.
- Schedule interviews only in public locations, such as public offices, libraries, fast-food restaurants, or other well-lit, high-traffic facilities.
- Train staff to take appropriate logistical precautions (such as having a clear set of directions and a map, a sufficient quantity of gas, accessible cell phones preprogrammed with key contact information and local numbers for assistance, and knowledge of public transportation options or how to arrange for taxicabs).
- Prepare field evaluators to watch out for their own safety. Teach them the kinds of situations to avoid (such as parking in isolated locations, walking a long distance in unfamiliar territory, and publicly displaying large amounts of cash or expensive jewelry).

Project Orientation. Unless the fieldwork is of very short duration or involves fewer than three field researchers, the team should prepare a document specifying procedures that field data collection staff are to follow. Such a document is helpful for both previsit training of the teams and for use as a reference while in the field.

The project field documents should review the following topics:

- The overall objectives of the evaluation and the specific purpose of the field visits.
- Human subject protections and data security, including the provisions of an institutional review board (IRB) plan if such a review was required.
- Item-by-item instructions for administering instruments, including definitions of terms used in the project.
- Advice on how to gain respondent cooperation and, when necessary, procedures appropriate to obtaining informed consent.
- Confidentiality requirements, including privacy during interviews.
- Procedures for conducting interviews, and circumstances under which interviews should be terminated and contingency plans for such emergencies should hold sway (e.g., if an interviewer perceives the respondent is

- under the influence of drugs or alcohol, or is showing signs of mental or physical distress).
- Other procedures for collecting data, such as carrying out structured observations, distributing or administering questionnaires, or auditing records.
 - Data storage, maintenance, and security procedures.
 - Quality control procedures, including instructions on when additional follow up is recommended for clarification, and how to edit field notes.
 - Administrative requirements, such as accounting and reporting procedures for dispensing incentives or submitting expense reports, obtaining reimbursement for travel, per diem rates, and so on.
 - Recommendations for managing time while on site.

Training. Training for field teams should cover all aspects of the field visits, including going over the instruments and all procedures in detail. Another important topic to include in the training is how to gain and maintain respondent cooperation. The level of cooperation secured will be partially dependent on the interviewer's ability to listen to the respondent, being aware of any sensitivities or anxieties the respondent might have, and responding appropriately to place the respondent at ease.

Before going into the field themselves or sending staff there, evaluators should consider the kinds of issues or resistance that respondents may raise. Training should incorporate answers to anticipated questions and should include having team members practice appropriate responses to likely situations or procedures to follow in the case of unforeseen events that could pose a threat to data collection. In large evaluations, these mock interviews are sometimes videotaped to provide immediate and forceful feedback to the interviewers.

Site Packets. Before the field trip, research staff should review the planned on-site procedures to assess their need for supplies and equipment, such as writing implements, notepads, electronic devices, and other office supplies that they should bring. We have found that two long (one hour) or three short interviews can be recorded in a notebook the size of a journalist's notebook (about six by eight inches). Staff should plan to take a few extra notebooks on each field visit to avoid running out. Similarly, if staff plan to use laptops to record information, preparation should include sufficiently long electric cords to charge devices during interviews and back-up paper and writing implements in the event of electronic failure.

If visits to several sites are linked or researchers anticipate collecting large amounts of bulky material, it may be desirable to bring along prepaid mailing labels and envelopes to send completed materials back to the home office. Because data, such as interview responses, may not be replaceable if lost, it is probably wise to use courier-type delivery services, which have sophisticated tracking capabilities, virtually guaranteeing that packages will not be lost. Electronically recorded information should be backed up each day, if not after each interview. Where confidentiality is a concern, arrangements should be made in advance to transmit these data using secure procedures (e.g., secure file transfer protocol or SFTP).

On-Site Procedures

Information on important on-site activities for which evaluators must be prepared should be included in the field visit protocol (procedures) developed before conducting the site visits.

Maintaining the Schedule and Interviewing Protocol. It is not possible to guarantee a particular response rate in advance. However, with appropriate planning and effort (such as careful scheduling, following guidelines for encouraging respondent cooperation, and having contingency plans for various situations), the team should be able to achieve a high interview completion rate, approaching 100 percent of those scheduled in advance. Professional demeanor and ability to conduct interviews without exhibiting judgment or excessive sympathy or emotion are particularly important.

If a respondent is reluctant to cooperate, field staff should attempt to convert reticence into cooperation. A first step is to ascertain whether the respondent has concerns or questions about the study or the interview that can be resolved, thus permitting the interview to proceed. For example, the timing may be inconvenient, in which case rescheduling the interview might resolve the impasse. In some cases, a rescheduled interview can be completed later in the visit or, if necessary, at a later time by telephone or videoconferencing. If this does not succeed, it may be best to allow another interviewer or the evaluation supervisor to attempt the interview at a future time.

A pilot test, or pretest, should be conducted in at least one site to try out all instruments and procedures. This will help to identify revisions or corrections that are needed. In most evaluations, the pilot site can also be part of the formal field evaluation because most of the same information will be collected.

Collecting and Recording Information. Field evaluators should be given materials that both help them to explain the purpose of the study to respondents and permit them to move efficiently through the planned interview. Each interview should begin with a brief introduction to the project. (An example is provided in Box 17.2.) Each respondent must understand the project and the purpose of the interview. Explaining the project takes only one or two minutes and is one of the most important parts of the interview. In these first few minutes, the interviewer should establish a rapport that places the respondent at ease.

Box 17.2. Sample Introduction and Statement of Confidentiality

"It is very important for us to learn how the Program for Family Independence operates here, problems or issues you have identified, and suggestions for how a program like this should be run. This is a new program, and it is essential to document its implementation and ongoing development.

"We need your cooperation to do this, since you know the most about welfare and employment and training problems and the problems that must be overcome. We are not employees of any state agency, nor are we auditors. We will be submitting reports to the state legislature, but in these reports, no one will be able to identify what any particular individual told us. We pledge confidentiality. The sources of our interview information will not be divulged to anyone else here in this office, city, or state. No names will be included in our reports.

"Do you have any questions before we begin?"

Immediately after the introduction, the interviewer should address confidentiality (also shown in Box 17.2). If the respondent's name is going to be included in a report, that must be explained at the time of the interview. If all information is to be confidential, meaning no names will be included in the report and no findings will be attributed to anyone by name, that needs to be explained. Confidential interviews are more likely than "public" interviews to produce rich detail—if the respondent understands the confidentiality pledge and believes that the interviewer will abide by the pledge. This is true even when the information being requested is not sensitive.

The evaluation should establish procedures for handling and storing the information collected in an interview, particularly if it is confidential. The procedures for maintaining confidentiality may range from not entering respondents' names into any databases that are constructed to devising systems of randomly generated identification numbers maintained in secured computer files.

After the introduction it is helpful to break the ice with an initial question designed to obtain background information on the respondent and ease into the interview. This can include asking the respondent his or her official job title, length of time employed with this agency, and what he or she did before this job. After this, the evaluator should move into the substance of the interview.

There are several ways to record the information from an interview: audio- or videotaping, taking notes by hand and transcribing the information into a notebook or directly onto interview forms, or using an electronic portable notebook or tablet, where notes can be written in longhand onto the electronic pad and later converted into a standard word processing format. Many evaluators now use laptop or tablet computers pre-loaded with data collection instruments, and type responses directly into such forms in appropriate places. Electronic pens are also available, allowing field teams to record notes that convert directly as a word processing document.

There are pros and cons associated with each approach, and the decision about which to use is generally a matter of personal preference. Subsequent analysis of information is generally accomplished more easily when interview notes (whether initially obtained through longhand notes or electronic methods) are transcribed into standard word processing following an established outline or topic format.

Daily Reviews. After the day's data collection is completed, evaluators should review the material gathered to add subject codes, respondent codes, site codes, or other explanatory details where needed. The material should be cleaned and clarified to be sure it is legible and meaningful to other members of the evaluation team. For example, only agreed-on abbreviations or those defined in the interviewer comments or other notes should be used.

Some evaluators choose to dictate each day's interview notes or reflections on other materials into a tape recorder. The tapes can later be transcribed for analysis or preparation of site reports. Some electronic pens also have audio recording capabilities in addition to electronic text conversion. The taping process also allows the evaluators opportunities to review the day's information carefully.

If there is more than one evaluator on site, the members of the evaluator team should briefly review their respective notes, data, and experiences to identify possible areas of inconsistency, issues that may have been missed totally (such as a question none of the respondents was able to answer well because it was outside all respondents' scope of responsibility), or areas that need further clarification or detail. The end-of-day debriefings afford valuable

exchanges of information that can be helpful later in the analysis. The team may want to record those sessions.

When an interview is finished, the respondent may express interest in knowing what the evaluators are finding. Similarly, when all data collection is completed at a site, an administrator may want to discuss with the evaluators the findings or conclusions they have drawn. Evaluators will naturally be thinking about preliminary findings before they leave the site, but they should not attempt to draw conclusions or make recommendations while in the field. Later phases of the evaluation (after the site visits) should be devoted to analysis. It is very tempting to provide immediate feedback but also extremely risky, so it should be avoided. Evaluators should be prepared to respond politely to such requests by explaining that they have accumulated a large amount of information and material that will have to be carefully reviewed and analyzed before it can be reported. This response may make the evaluator slightly uncomfortable, but it is much better than realizing later that he or she has given a program official partial or incorrect findings. To soften the exchange, evaluators can briefly review with the site principals what the researchers plan going forward (e.g., when follow-up visits will occur, when a draft report is anticipated, whether sites will be given the opportunity to review and comment on draft materials, and when and how final reporting is expected to be made public).

Data Maintenance and Analysis

Once field data have been amassed, they can be used to generate several types of summaries, such as quantitative frequencies, trends, contingencies, and intensities. In addition, qualitative information can provide rich anecdotal evidence.

Most important when observation or semi-structured interviews are used, evaluators need to decide how to systematically summarize the large quantity of information collected. Analyzing qualitative data is roughly equivalent to performing analysis of more structured data collection methods such as surveys, in which the documents under scrutiny are the records of interview responses. Such analysis involves organizing the data into relevant sets of content or issue categories or topics and sets of response alternatives for each content or issue category.

Even when data sets are derived from semi-structured, open-ended instruments, evaluators should identify preliminary categories of possible responses prior to data collection. This structure provides guidelines that help to orient

the data collection efforts. For example, anticipating certain themes and possible responses can help field staff determine whether observations, interviews, or record extractions are achieving the evaluation objectives or whether evaluators need to probe further. Usually, preliminary categories or topics can be proposed during the fieldwork planning stage, based on the evaluation questions or hypotheses. Often the range of response alternatives can also be anticipated. However, these predetermined coding possibilities should be viewed flexibly because new themes and insights are likely to emerge during the data collection, or some anticipated topics and responses may never materialize.

One approach to data organizing is to have field evaluators sort the information for each identified research topic or category, using the response alternatives postulated prior to actual data collection. The coding scheme is then finalized in light of feedback from field staff about categories or response alternatives they have identified that do not fit the data. Adjustments can be made by expanding or collapsing the initial choices of topics or responses. As analysts sort the data, they can also flag any anecdotes or quotations that might enrich the final report.

A more rigorous (but also more costly) approach to analyzing qualitative interview notes, if resources are sufficient, is to divide the analysts into two teams to review a sample of observations, interviews, or extracted records independently. Each team develops a set of content categories and response alternatives for each category based on the data sources. The teams convene as a single group to discuss and merge the categories and responses. Once consensus is achieved, the data can be split into subsamples. Each team receives one subsample, which team members code using the agreed-on scheme; then the teams exchange samples and repeat the categorization process. This approach tests both intracoder reliability (the degree of consistency with which a coder interprets similar responses) and intercoder reliability (the degree of consistency in interpretation among different coders).

If consistency in coding is unacceptably low, there are several options for improving reliability:

- Categories can be tightened and redefined to reduce the chances for mis-coding.
- A training session can be held to increase intercoder reliability by making coders more familiar with the categorization system.
- Instead of having analysts code every item for a series of observations or interviews, each analyst can be assigned responsibility for coding the same set of questions or topics for all observations or interviews, thus becoming

the coding specialist for specific items, and thereby increasing intracoding reliability.

The data maintenance approaches described above refer to situations when the information collected in the field is manually sorted and coded. Researchers can use electronic tools instead of, or in combination with, manual procedures. Technology is rapidly developing, and there are various creative electronic tools for recording, coding, sorting, and analyzing field data and notes. For instance, there are many content analysis software programs that can be used to generate lists of unique words or phrases in one or more documents, as well as enumerate the frequency of occurrence. These and other programs have streamlined qualitative data analyses.

Conclusion

The fieldwork portion of an evaluation provides an opportunity to collect rich detail that can augment more quantitative data that are included in the evaluation. Too often, fieldwork is approached in an informal or haphazard manner that results in massive amounts of notes and other information that cannot be easily analyzed. Evaluators should pay careful attention to developing fieldwork procedures, designing fieldwork data collection instruments, and preparing plans for managing and analyzing the information collected. Carefully implemented, fieldwork data collection can produce valid and credible information that cannot be obtained from other sources.

References

- Abbott, A. "Sequence Analysis: New Methods for Old Ideas." *Annual Review of Sociology*, 1995, 21, 93-113.
- Bellotti, J., Rosenberg, L., Sattar, S., Esposito, A. M., and Ziegler, J. "Reinvesting in America's Youth: Lessons from the 2009 Recovery Act Summer Youth Employment Initiative." Princeton, NJ: Mathematica Policy Research, 2010.
- Hernandez, M. "Using Logic Models and Program Theory to Build Outcome Accountability." *Education and Treatment of Children*, 2000, 23, 24-40.
- Holcomb, P., and Nightingale, D. S. *Evaluation of the Western Interstate Child Support Enforcement Clearinghouse*. Washington, DC: Urban Institute, 1989.
- Mitchell, J. J., Chadwin, M. L., and Nightingale, D. S. *Implementing Welfare Employment Programs: An Institutional Analysis of the Work Incentive Program*. Washington, DC: Urban Institute, 1979.
- Nightingale, D. S., and Ferry, D. *Assessment of the Joint Federal-State Performance Improvement Projects in the Work Incentive Program*. Washington, DC: Urban Institute, 1982.