

Using Case Studies to do Program Evaluation

Evaluation of any kind is designed to document what happened in a program.

Evaluation should show: 1) what actually occurred,
2) whether it had an impact, expected or unexpected, and
3) what links exist between a program and its observed impacts.

First, knowing what actually occurred in the course of the program means knowing who did what to whom. Information on what happened is captured in what are known as process measures, outputs or, sometimes, widget counting. These measures are not sufficient to determine whether or not the program had an impact, but without them it isn't clear whether the program was carried out according to plan. Evaluators need evidence that the program was delivered in order to make responsible statements about causation.

Second, the evaluator needs to know the impacts of the program, both those the program is expected to achieve and also unexpected positive or negative impacts. The expected impacts are codified in statements of program objectives. An evaluator often takes baseline measurements to document the situation before an intervention, and then takes the same measurements after an intervention to assess impact. However, this technique only works for expected impacts. The unintended impacts are more challenging to document, because there are usually no baseline measures for them. Yet, identifying unintended impacts can be important in understanding all of the impacts of a program.

Third, the evaluator needs to make responsible judgments about the links between a program intervention and its observed impacts. While a certain amount of conjecture enters into making these links, increased knowledge about the program as it was delivered, and its impacts, leads to stronger conjecture.

Most traditional evaluation designs use quantitative measures, collected over a sample of the population, to document these three stages. However, there are times when this sort of evaluation design does not work as effectively as a case study evaluation. This guide is designed to help evaluators assess whether or not a case study is a useful evaluation tool for a given project, and if so, this guide explains how to do a good case study evaluation. A list of resources is included in Appendix A. Like any other evaluation design, the case study should suit the project to which it is applied and must be well executed for maximum benefit. The guidelines found in the document, *Tell Your Story: Guidelines for Preparing an Evaluation Report*¹ can aid evaluators in preparing case study evaluation reports.

This guide will help evaluators assess whether to use a case study evaluation approach and how to do a case study

¹Albright, A., Howard-Pitney, B., Roberts, S., and Zicarelli, J. (1998). [Tell Your Story: Guidelines for Preparing an Evaluation Report](#). Sacramento, CA: California Department of Health Services.

Table of Contents

Introduction	1
Using a Case Study as an Evaluation Tool	3
When to Use a Case Study	4
The Program is Unique	4
An Existing Program in a Different Setting.....	4
A Unique Outcome.....	4
An Unpredictable Environment	5
The Strengths of the Case Study	5
How to Do a Case Study	6
Unit Selection	6
Sampling Techniques.....	6
Random Samples.....	6
Purposive Samples.....	6
Convenience Samples.....	7
Data Collection.....	7
Interviews.....	7
Structured Questionnaires	8
Semi-Structured Interviews.....	8
Completely Unstructured Interviews	8
Concluding Thoughts on Interview Protocols.....	9
Tape Recording Interviews	9
Selecting People to Interview	9
An Example: Interviewing in Practice.....	10
Observation	12
Document Reviews	12
Data Analysis and Interpretation	12
Two Facets of Data Collection and Analysis.....	12
Analyzing the Data	13
Showing a Plausible Link between the Outcome and the Intervention.....	14
Sources of Error	16
Inaccuracies	16
Bias	16
Competing Explanations	16
Selection Mistakes	16
Presentation	17
Conclusion.....	17

Using a case study as an evaluation tool

The purpose of a case study is to study intensely one set (or unit) of something—programs, cities, counties, worksites—as a distinct whole. What does this mean? For a program designed to encourage bars to observe the smokefree bar law, an evaluation must document the program’s impact on the bars and on the behavior of people in the bars. In a non-case study design, one might decide to observe a series of randomly selected bars to see whether bartenders take some action to enforce the smokefree bar law when customers begin to smoke. This style of evaluation entails collecting data on bartender behavior from a random sample of bars large enough to be representative of the entire population of bars from which you sampled.

In contrast, a case study design focuses on a hand-picked set of bars (sometimes even just one bar). Before the program begins, the evaluator spends time in the bar(s), observing behavior and talking with people. As the program progresses, the evaluator continues to make observations and to interview the owners, managers, employees, and customers. She might observe the bars at various times of the day to monitor compliance with other smokefree rules, such as the absence of ashtrays. At the completion of the program, the case study reveals in depth the experience of specific bars in implementing the new law, and the impact the program had on its efforts. Did the program either encourage or discourage compliance? Did new signage go up, and did bartenders begin to encourage compliance? Or did something completely unrelated to the program happen to change behavior?

For example, did a bartender relapse during a quit attempt and resume smoking, thus encouraging others to smoke? Did, on the other hand, a favorite bartender have a heart attack, which made the customers more sensitive to smoking behavior?

This kind of rich detail lets evaluators assess programs in a way that several data elements across a large variety of cases cannot. In a case study, note that some of the data collected might be quantitative, such as the number of instances of compliance at various times of the day. Case studies do not necessarily use qualitative data only. Overall, case studies are considered to be a qualitative technique, but they can contain quantitative information. However, the overall goal of a case study, which is to understand a select subset as a distinct whole in its particular context, distinguishes the case study from other designs.

What one gains in richness by doing a case study evaluation, one loses in the breadth of generalizations about overall compliance. Put another way, a case study reveals a lot about the process and outcome at certain sites, and the ways in which these interrelate. It reveals less about a program’s overall impact. One way to offset the lack of breadth in a single case study is to do multiple case studies and to compare the findings. For example, an evaluator could do complete profiles of several sets of bars, compare their implementations of the smokefree bar law, and look at the similarities and differences in implementation. This comparative study begins to clarify the impacts that your program either had or did not have, providing useful information for program revisions.

The purpose of a case study is to study intensely one set (or unit) of something—programs, cities, counties, worksites—as a distinct whole.

Case studies do not necessarily use qualitative data only.

When to use a case study

So when is a case study an appropriate evaluation tool? A case study **SHOULD NOT** be done simply because no other evaluation design seems to work. Case studies take considerable time and energy to do well. They should only be used for program evaluation if they are the best way to answer the “what happened?” question. A case study is particularly useful for evaluating programs when programs are unique, when an established program is implemented in a new setting, when a unique outcome warrants further investigation, or when a program occurs in an unpredictable environment.

The Program is Unique

If a program is highly innovative, such as a coalition’s effort to help a sister city in Asia work on tobacco control, then it may be extremely difficult to predict the program’s positive and negative impacts. However, it is still necessary to document those impacts systematically, and to consider whether those impacts resulted from the program. In addition, the rich detail of a case study provides good information about the design of a program and the context in which it is delivered, thus allowing others to determine its appropriateness for their areas.

In evaluating the sister city relationship, as with most case study designs for unique programs, the case study should begin in the program’s planning phase. The evaluator should observe as many meetings as possible, and interview participants as they plan and implement activities. Interviews conducted after the program is complete will depend on people’s recollections of events. It is better to interview people as the program takes place.

Case studies take considerable time and energy to do well.

Use case studies if they are the best way to answer the ‘what happened?’ question.

An Existing Program in a Different Setting

A case study can also be a useful evaluation tool when the project involves implementing an existing program in a new setting. For example, if a cessation program developed for smoking teenagers is being used for the first time with those who use chewing tobacco, then a case study evaluation design could best fit the chewing tobacco program. This design allows for complete documentation of what appeared to transfer easily and what did not. A standard pre-test and post-test design incorporated into the case study might help determine whether chewers are as likely to benefit from the program as smokers, but the point of the case study evaluation is to understand the details of how the program worked in a new setting. A case study evaluation design would also be useful for a cessation program designed for voluntary attendees, but used instead with students ordered to attend for smoking on campus. How does the material and delivery change with a new population? As with case studies that evaluate unique programs, case studies that evaluate an existing program in a new setting should begin in the planning stages of the program.

A Unique Outcome

In the course of doing a purely quantitative, non-case study evaluation, an outlier may occur (an instance in which the outcome for one member of the population differs greatly from the outcomes for the rest of the population). It may be that what happened in that one instance is sufficiently different to warrant an in-depth case study.

For example, for a program to educate parents about the need for smokefree homes, an evaluator begins with a non-case study evaluation to assess the program’s impact. Before starting the program, she conducts a random sample telephone survey of two districts exposed to the program and of a non-exposed (control) district to determine

the current rate of smokefree homes. Every 6 months for the next 3 years, she conducts a random sample telephone survey of the exposed and control districts to track the rate of change over time. One of the data elements collected is the school the children of the household attend, or some other indicator of home location. If one geographic area or one school catchment area is significantly better or significantly worse than the others, then the evaluator might do a case study of that area. An interesting reason for the variation could point to future intervention strategies. For example, if only one geographic area has a community-based clinic that provides tobacco education, and that area did better than the rest of the city, then the program might promote tobacco education in other clinics in future efforts.

Unlike case studies that evaluate unique programs or established programs in new settings, case studies that evaluate unique outcomes must take place after the program has been implemented. The evaluator is likely to have more documents to review in this style of case study, but less likely to be able to observe behavior. Most of your interviews will depend on recollections of events.

An Unpredictable Environment

When the environment is complex and turbulent, the achievement of a pre-established goal may be difficult to predict, or may only be achievable with unacceptably large negative side effects. A case study evaluation for a program implemented in a turbulent environment should begin when program planning begins. A case study evaluation allows you to create a full, complex picture of what occurs in such environments. For example, ordinance work is pursued in political arenas, some of which are highly volatile. It is only somewhat interesting from an evaluation standpoint to record whether or not an ordinance passed. This outcome is, of course,

readily measurable and easy to ascertain. However, the more interesting evaluative information would include a description of the obstacles that faced tobacco control advocates and a description of whether and how the advocates overcame them. Was the tobacco industry present? What suspicious behavior did the evaluator see? What information or tactics were useful with the political body? Could another jurisdiction use similar ones? A case study evaluation helps clarify the whole process and how it may have affected the final outcome.

The Strengths of the Case Study

Traditional evaluation designs generally assume that program implementation follows a rational, predictable, and measurable path. Some programs fit this profile better than others. One of the strengths of a case study evaluation is that it is not based on that assumption, and instead treats program implementation as a series of events, each of which calls for new strategies that may end up changing outcomes. A case study evaluation allows greater latitude in seeking out and assessing program impacts.

In Figure 1, if more 'yes' than 'no' responses to the questions apply to the program, case study evaluation might be useful. Case method evaluation is time-consuming, however, and only advisable when it is very clear that knowledge can be gained from studying one unit in detail.

Case study evaluation allows you to create a full, complex picture of what occurred.

Figure 1: When to Use Case Methods

	Yes	No
Is the projected program unique?	<input type="checkbox"/>	<input type="checkbox"/>
Is it premature to establish impact measures?	<input type="checkbox"/>	<input type="checkbox"/>
Are the projected program impacts too impractical or too difficult to measure?	<input type="checkbox"/>	<input type="checkbox"/>
Is there unexplained variation in the program impacts?	<input type="checkbox"/>	<input type="checkbox"/>
Will understanding the program implementation in detail help anybody design future programs?	<input type="checkbox"/>	<input type="checkbox"/>

How to do a case study

A completed case study report will have detailed descriptions of what happened, and the context in which it occurred, from multiple viewpoints. The report will feature a factual recounting as well as an interpretation of events. This section of the manual explains the necessary steps for arriving at a finished product.

Unit Selection

In designing a case study evaluation, one must decide how to select the case to be studied. Is it desirable to track how a particular group is implementing a program? Is it beneficial to understand one organization's behavior in depth? If so, by what criteria does one select a group or organization? Is it important to study how the program operates in several different kinds of sites, necessitating a multi-case design? On what dimensions should the sites vary? These considerations determine unit or set selection.

Sampling Techniques

There are three main sampling techniques used in case study evaluation: random, purposive, and convenience.

Random Samples

Random samples require the creation of a complete list of all the units in a population from which units of that population are selected randomly to study. For example, from a complete county list of establishments with a license to sell alcoholic beverages by the glass, one might randomly select a sample to study over time. Random or comprehensive sampling of the population of interest characterizes most non-case study designs. It is not often used in case study evaluations.

Purposive Samples

Case study evaluations almost always use purposive samples. Purposive samples are used when the evaluator is studying a particular phenomenon and wants to ensure examples of it show up in the study. The object of a case study evaluation is not to find out how often something occurs in a population, but rather what occurred, why it occurred, and what relationship exists among observed events. Purposive sampling is frequently used in case study evaluations because, in order to study the reasons something either did or did not happen, an evaluator must be sure that the units in the sample have the potential to reveal those reasons.

For example, if an evaluator wants to use case study methods to evaluate the successes in getting bars to comply with a smokefree law, he might want to do a multiple case design, in which he compares bars that complied with those that did not. He could pick five bars that went smokefree easily and compare them to five bars that did not. He then studies each bar—its patrons, owner, location, distributors, etc.—and draws a complete picture of its smokefree effort. By comparing the successful bars with the unsuccessful ones, he begins to understand how various pieces interact to get the observed outcome (smoking or not). He is using a purposive sample. Using the criteria smokefree versus non-smokefree, he selects bars to study.

If an evaluator instead took a random sample of bars to study, she could end up with 10 bars with good implementation, which means she lacks the contrast that yields helpful information. This risk is particularly high if the majority of bars are smokefree, and she wants to know more about the ones that are not. If the majority are smokefree, then a random sample of all the bars in her area is likely to yield smokefree bars. She could overcome

Case study evaluations almost always use purposive samples

this difficulty by dividing the bars into smokefree versus not smokefree, and then pulling a random sample. Similarly, she could pull a random sample of all bars, and then do a screening visit to them in order of selection, studying the first five that were smokefree and the first five that were not. In either case, she still needs to establish groups based on some criteria, in this case, being smokefree or not.

Convenience Samples

Convenience samples are drawn when other sampling is not practical and one can get reasonably good information from units that are easy to locate. For example, if a neighbor owns a bar and is willing to cooperate with a case study, then his or her bar could be included in the study, provided it will yield information, and that studying a friend's bar will not affect the evaluation. Better access will yield better information for the case study; access to people and places is extremely important in doing case study evaluation. Neither a random sample nor a purposive sample yields a better case study if little is learned about the bars because they fail to cooperate. The risk in taking a convenience sample is similar to the risk in taking a random sample for a case study. The sample may not include units of study that vary on the important dimensions. There is also a risk that bars that volunteer to be studied differ systematically from those that do not. Furthermore, there is a risk that bars identified through friendship networks may be quite different from all other bars in the area, which means much may be learned about how bars an evaluator might frequent will respond, but little about bars that serve other groups of patrons. This is still valuable information, but the limits on it should be made clear when the interpretation of the data begins.

Data Collection

Data collection in case study evaluation is designed to answer the classic journalism questions: who, what, when, where, and why. Specifically, the case study evaluator needs to know:

1. Who was involved in the program?
2. What did they do, in terms of activities?
3. In what context were they working: political, organizational, cultural, etc.?
4. When did the program activities take place?
5. Where did the activities take place?
6. Why did participants do what they did?
7. What, if anything, about the actions taken caused the observed changes to take place (if indeed there were changes)?

But how are the answers to these questions found? The chief sources of information in doing case study evaluation are interviews (including focus groups), observations, and documents. With multiple sources of data, one generally can draw a more complete picture of what occurred and why.

Interviews

Although a case study evaluation can be done without focus groups, documents, and observations, it is rare that it is done without interviews. For this reason, interviews are the foundation of case study evaluation. Interviews are the path to understanding both what happened from the perspective of those involved and how they reacted to it.

Interview protocols can range from highly structured questionnaires, in which the questions and range of answers are specified in advance, to nondirective conversation, in which the respondent has considerable latitude to talk about what he or she chooses. Between these extremes are mixed forms in which some specific

Interviews are the foundation of case study evaluation.

questions are asked of all respondents, but there is also room for the collection of less structured information. The decision to use one form or the other depends on what the evaluator is trying to learn.

Structured Questionnaires

Highly structured questionnaires are not often used in case study evaluation. These questionnaires generally are designed to capture standardized information from all respondents, and are used for projects such as the California Tobacco Survey. In these surveys, the interviewer assumes that she or he knows enough about what has happened to know all of the appropriate questions and all of the possible answers. A tightly structured questionnaire is almost impossible to design in situations that call for case study evaluation since the interviewer does not know enough about what occurred to know all of the appropriate questions and answers. The attempt to administer such a structured survey is likely to prove frustrating for both the interviewer and the respondent.

The interviewer may, however, wish to use a short version of such a structured questionnaire to collect basic demographic information about respondents, if it is relevant to the study, or to collect standardized information based on observations. For example, to study the implementation of a smokefree project at a local teen center, one might want, at regular intervals, to count the number of cigarette butts in the outdoor ashtray as part of a larger study. This type of data is collected on a standardized form.

Semi-Structured Interviews

In general, case studies use less structured interview protocols and usually include primarily open-ended questions. The interviewer has a better chance of learning about the perceptions and experiences of those being studied by using open-ended questions. It is unusual, however, for a case study evaluation to include a completely unstructured interview protocol. Most case study evaluations

include an interview protocol that specifies in advance the general topics of interest. The use of a protocol insures that similar information is collected from all respondents. If the results of several case studies are to be compared, similar information must be collected across the cases.

Sometimes the topics in an interview protocol are listed in the form of questions, but they need not be. It is not necessary for the same wording to be used with respondents, or for the topics to be covered in the same order with each respondent. The object of the interview is for the interviewer to know, by the end of the interview, how the respondent perceived what happened. The underlying assumption is that respondents experience the world in unique ways, and the object of the interview is not to standardize their experiences, but rather to understand them.

Completely Unstructured Interviews

In contrast to semi-structured interviews, completely unstructured interviews allow maximum flexibility in collecting the experiences of the respondent. This kind of interview is helpful in case study evaluation when the evaluator knows little about what happened or about the organizations involved. For example, if an outside evaluator is trying to do a case study evaluation for a community-based organization, the evaluator might begin interviews by asking people to describe the organization and how it came to be positioned in the community, encouraging respondents to tell stories about the organization. The evaluator might also want to know the history of tobacco control work done by the organization. While this process does not necessarily result in the collection of the same categories of information from all participants, it does give respondents a chance to express themselves. After conducting these unstructured interviews to learn about the organization, the interviewer might then review paper documentation, such as a grant proposal, to learn about the specific project being evaluated.

The use of protocols insures that you get some comparable data from each respondent.

If more information is needed about the program, the evaluator might know enough at this stage to create a semi-structured interview protocol for the rest of the evaluation. The early unstructured interviews might establish a program baseline, or constitute a pilot study for a larger case study evaluation. However, the unstructured interviews alone might yield sufficient information to explain what happened.

Concluding Thoughts on Interview Protocols

A good interview protocol shapes the interview without constraining it. That is, the respondent will feel that the interviewer did her homework and understands the key issues, but will also feel welcome to volunteer information. The use of semi-structured protocols produces comparable data from all of the respondents, and also gives them room to discuss things they perceive as unique. With some experience, the interviewer might find it useful to converse with the respondent without using the protocol, and then check the protocol part way through the interview to see what areas remain to be covered.

The bottom line in interviewing is that the interviewer must develop the ability to establish rapport with the people being studied, to ask questions that allow people to express themselves, and to listen carefully to the answers given. The interviewer who dominates the conversation will learn very little.

Tape Recording Interviews

Whenever possible, interviews should be tape recorded. Note taking can distract the interviewer from engaging the respondent in conversation. In addition, it is hard to get good quotes without tape recording. The use of quotes in a case study write-up helps demonstrate that the evaluator has captured the respondents' perceptions and feelings accurately. Paraphrasing provides weaker evidence.

It is always preferable to transcribe tapes fully, but this is very time consuming; a one-hour interview can take four to five hours, or even longer, to transcribe. Sharon Merriam, in her book *Qualitative Research and Case Study Applications in Education*, suggests a possible alternative to full transcription, which she calls an interview log. In an interview log, the evaluator listens to the tape and takes notes on the respondents' comments. Some sentences or whole paragraphs are transcribed verbatim so they are accurately quoted in the final report. The tape counter is used to mark the quotes' positions in the interview. If there is no time or money to do full transcription, such logs are an improvement over sole reliance on note taking.

Some notes should be taken during an interview, however, in case the tape recorder fails. Immediately following an interview, the tape should be checked to see whether it recorded. If it did not record or if the sound is bad, the interviewer should transcribe her notes immediately, capturing other details she remembers. The longer she waits to do this, the more she will lose. The following steps will ensure the tape recorder works: 1) checking the batteries before an interview, 2) testing the tape before the interview, 3) turning off voice activated recording, if the tape recorder has it (otherwise the first few words of many sentences will be lost), 4) if a table microphone is used, making sure that it is on and that it has good batteries, and 5) trying to find a quiet place for the interview. While it is sometimes easiest to establish rapport over a cup of coffee in a public place, a lot of background noise might get recorded.

Selecting People to Interview

Generally, in interviewing for case studies, the interviewer should first seek out those who know a lot about the program and are willing to be reflective about it. These people are sometimes called key informants, and can help set the program in its context and provide the names of others who would be valuable to interview.

Advice about other people to interview is useful because, unless the program is quite small and involves few participants, decisions must be made about who will or will not be interviewed. The same sampling issues that arise in unit selection also arise in selecting people for interviews. In general, as with unit selection, one should think about the purpose of the interviews and let the purpose determine the selection of respondents. Because the case study is done for purposes of evaluation, it's usually appropriate to include some of the people for whom the program was developed.

An Example: Interviewing in Practice

Perhaps a program involves youth in an effort to convince merchants to voluntarily pull tobacco signage from their store windows. One measure of success is a reduction in the amount of signage. However, another purpose of the program could be to effect a change in community norms regarding the appropriateness of tobacco advertising. The comparative case method might reveal in some detail reasons for merchants' action, and the extent to which those actions reflect and affect a community norm change.

To determine how a particular set of stores reacts to a program, a subset of those targeted by the program should be selected. Each selected store is a case within the study of the overall program. The evaluator might pick the stores by location, randomly, or by other relevant criteria. At an early stage, he might want to interview program personnel about their plans and goals, interview the owner, observe the social context of the store, and collect any other data that would help him understand that store and its neighborhood. Throughout the program, the evaluator should keep checking back. As the program progresses through its planning and implementation stages, he should continue to interview the program implementers, including the youth who are involved.

Once quantitative data on the impact of the program are available, the evaluator should note which stores made changes and which did not. If nearly all of the stores being followed made changes, or nearly all did not, some stores may be added to the study that are different from the ones in it. One way to determine which stores are different is to mark on a map those that changed and those that did not, so that any obvious geographical clusters emerge. The number of additional stores needed for a comparative case study effort should become clear. At the stage of adding stores to study, the evaluator might confer with the program implementers to get their perspective on stores that could be rich sources of information, either because they were cooperative or because they were not. Clearly, for the added stores, the study will be retrospective.

How should the interviews required in this evaluation be conducted? It is logical to begin with those involved in program implementation, in order to learn details about the program. It is best to interview people at a time that is convenient for them, because it is important to accommodate someone kind enough to contribute his time. Doing so will result in a better interview. In the early interviews, an unstructured interview protocol will draw as much as possible from those most involved in the program. At the beginning of each interview, the interviewer should emphasize that he wants to have a conversation about the program. This type of opening helps people to understand the unstructured format of the interview.

The interviewer should start by asking some general, open-ended questions that are easy for the respondent to answer, and that will get him or her in the habit of talking, such as, "Tell me about how you put together your program?" This question allows respondents to talk about something they know, and lets them know that the information they volunteer is important. A series of follow-up questions might follow, designed to elicit exactly

what was done at each location. These follow-up questions are called probes, and are designed to clarify what the respondent has done.

Toward the end of program implementation, the interview should elicit information about the quality and utility of the program. However, questions of this kind should follow some basic descriptive questions. By asking evaluative questions last, one gains enough time to establish rapport with the respondent before soliciting opinions. Most people are more comfortable volunteering opinions after having talked for a while. The interviewer should end by asking whom else she should interview, which is called snowball sampling. All or most of the people who can help render a complete picture of what occurred should be included.

Next, merchants should be interviewed. At this stage a semi-structured interview could help standardize the information collected across stores. Again, factual questions should come first, and opinion questions later. For example, in the later rounds of interviewing, one might start by asking merchants to describe how the youth approached them, then discussing who was involved in the decision to keep or remove the signage. A discussion about the reasons for the decision should follow. A direct question (e.g., Does this reflect a norm change in the community?) is unlikely to be effective. Questions about the kinds of pressures the merchants feel about signage, and about their stores' roles in the community, elicit more information. The interviewer should ask about customers' reactions to the signs. For those who removed the signs, it could be helpful to ask what would cause them to return the signs to the window.

One might also wish to interview a convenience sample of store customers, varying the time of day that data is collected. At this stage a short, highly structured questionnaire might best elicit their reactions to the change (or lack thereof).

“Do’s” and “Don’ts” of Interviewing

Do

1. Listen carefully and follow up. If you don't understand someone's point, or a word he uses, ask. If someone makes an interesting point, ask him to tell you more about it. In short, be interested. These skills are those of a good conversationalist. In fact, unstructured/semi-structured interviewing and good conversation have a lot in common.
2. Allow time for the interview. As soon as you become rushed, you lose information.
3. Summarize what you think you've heard. One useful way to keep an interview moving forward is to summarize the respondent's comments and to ask if you understood them correctly. This technique not only confirms your understanding but also lets the respondent know that you are paying attention.
4. Tape the interview, if possible. Even if you do not have the time or resources to transcribe a tape, you still learn more from listening to a tape than from taking detailed notes during the interview. Take some notes, however, in case the tape does not work.

Don't

1. Don't ask leading questions. You want to know people's opinions. You don't want to feed them yours. Sometimes it is appropriate to solicit these by playing devil's advocate, but use of this technique is best left to follow-up questions after someone has expressed an opinion.
2. Don't ask yes/no questions. Any time you ask a question that can be answered with a yes or a no, you are prompting people to give short answers. If you want people to give more descriptive responses, avoid questions that begin with can, will, did, or anything else that calls for a yes or no. Use instead words such as why, how, or tell me.

Interviewing and good conversation have a lot in common.

Observation

In case study evaluations, interviews are often complemented with observation and/or an analysis of documents. Both observation and document reviews are often useful supplements to interviews, and in fact sometimes turn out to be more important than interviews, depending on the nature of the program. Observation allows for more direct experience than an interview does, and can help clarify the context in which the program is implemented. In doing observations, one should maintain field notes, in which factual statements of what occurred, relevant quotes, and other potentially useful information can be recorded.

Observation allows you to see and experience directly what you otherwise would try to learn about in an interview.

For example, to evaluate the success or failure of an effort to pass an ordinance, one must spend time observing the political process: attending hearings, keeping track of who testifies, and noting the body language of those listening. When does a particular politician leave the room, and who does he or she speak with in the hallway? Interviews supplement observations, but observations should reveal some of the political dynamics surrounding a public action.

Document Reviews

In case study evaluations, paper trails can guide the development of the interview protocol, confirm comments made by respondents, or provide a second observation of an event. They can also shed additional light on the outcome of an event. The challenge for an evaluator is to validate what is written on paper, as paper trails can be used to obscure events as well as to codify and clarify them. For example, press releases often reflect what an organization wants to be true, but maybe not what its insiders would say really is true.

Be extremely conscious about challenging your pet theories.

Sometimes, however, documents are invaluable. For example, if one of the measures of the effectiveness of the program is the amount and tone of the media coverage, then newspaper stories are important. In trying to understand an organization's history on tobacco control, one should review documents such as past grant proposals, reports, and other documentation, rather than relying solely on the recollections of those involved. An evaluator should find out what is already written down. At a minimum, this effort produces background information on which to base interviews.

Data Analysis and Interpretation

Two Facets of Data Collection and Analysis

In conducting a case study, the evaluator often acts as both the chief data collector and the chief data analyst. This dual role presents both hazards and opportunities. The hazards are that the evaluator, like all human beings, enters into a situation with a set of beliefs and preconceived notions. As more data is collected, and the evaluator starts analyzing it, even informally, he or she may become more committed to these beliefs. The challenge for the evaluator is to prove himself or herself wrong by finding evidence that disconfirms a pet theory.

Challenging pet theories can be hard. For example, if an evaluator believes that a city councilor voted against a vending machine ordinance because of tobacco industry influence, she needs to spend time trying to prove that there might be another motive. In the end, she might still believe that the tobacco industry was influential, but she must seek out evidence that she might be wrong.

The opportunity in being both data collector and analyst is that, with the right amount of flexibility, the evaluator can respond to new avenues of inquiry as they are presented in the course of the evaluation. The method for doing a case study cannot be thoroughly planned, because as data collection unfolds and analysis begins, the evaluator becomes aware of new paths that are worth pursuing. It is often impossible to know in advance all the necessary questions, or all of the people with whom it would be useful to speak. Thus, good case study evaluation requires sensitivity to the environment and to people under study. This sensitivity allows the evaluator to make quick decisions, and take advantage of new opportunities. If an evaluator is comfortable only with a research protocol specified fully in advance, and executed with minimal adjustment, then the evaluator might not be comfortable with case study evaluation.

Analyzing the Data

The goal of the case study evaluation is to collect and present data from multiple sources in sufficient detail that a critical audience believes the story that is told. The test is plausibility, but a computer does not do plausibility tests. Judgment is necessary, as well as critical review by others.

One of the key contributors to conducting a plausible case study evaluation is to start data analysis and data collection concurrently. Analysis begins with the first document review, the first interview, or the first observation. The evaluator starts to build a theory about what is going on, and

then begins to think about how to confirm that theory. Part of confirming the theory is searching diligently for evidence that the pet theory is wrong. What this means is that one must decide what evidence would disconfirm the pet theory, and then try to find that evidence. If the evaluator fails to find it, then the theory and case study are strengthened.

As more information accumulates, the first theory can continue to be refined. Points of agreement and disagreement among program participants, interesting recurring themes, and categories of insights will emerge. Writing should start at the beginning and continue throughout. The writing focuses the evaluator's energies on identifying the points, themes, and insights that define the remainder of the study and the final report. It also helps one to identify where believability wanes, and suggests what additional information might be useful. Writing furthermore speeds the production of a final report.

How does this work in practice? If one is evaluating a sister city relationship, he might begin by reviewing coalition minutes to see what happened during the project from the moment it was conceived forward. Perhaps it comes to light in a review of the minutes that the coalition and the sister city decided to try to pass a ban on cigarette sampling in the sister city. As part of the case study, the evaluator must document the steps taken to achieve this ban. It turns out that the ban moved part way through the city government. Armed with this information, the evaluator decides to interview those most involved in working with the sister city to see what happened.

Present data in sufficient detail, from multiple sources, so that a critical audience would believe the story that is being told.

Start your data analysis as soon as you start collecting data.

Create a clear record of the methods used to perform the data collection and analysis.

If the first two people interviewed mention the key importance of having a bilingual coalition member involved in the project, four things should be done. First, a note should be added on interview protocol to remind the interviewer to ask others about it. Second, a section should be created in the final report for a discussion of this finding. It may drop out later, but one should begin early to look for patterns. This technique results in quicker production of an evaluation report, and improves data management. Third, it should be noted where on the interview record (either the transcript or the tape) this discussion occurred. This marking makes it easy to find the comment later. Fourth, a master list of the themes that seem to be important should be made, so that one continues to look for them in other interviews, observations, and documents. The final two steps in this process are sometimes called coding, and coding gets more complex, and more necessary, the larger the study.

In some really complex studies, computer programs are used to assist this effort. For a small evaluation of a minor program, a specialized computer analysis program may be more trouble than it is worth. The coding process itself, however, will maintain organization. A good word processing program can help with data management.²

Showing a Plausible Link between the Outcome and the Intervention

The final step in an evaluation is to make responsible judgments about the links between the program as it was delivered and its observed impacts, or lack of observed impacts. This step is fundamentally a matter of interpretation, and on this point rational people sometimes disagree. One of the important features of interpretation is acknowledging uncertainty, because in any method of evaluation some exists. In any analysis, one needs to let others know where interpretations are being made and the degree of confidence one places in them.

In experimental or quasi-experimental designs, the methods for acknowledging uncertainty are highly stylized. By contrast, qualitative research may include some descriptive statistics, but probably no tests of statistical significance. In drawing conclusions from case studies, an important step for evaluative purposes, the evaluator should draw conclusions so that a reasonably skeptical observer will believe them. (Note: The tobacco industry will never be convinced, so don't use it as a standard.)

The strongest conclusions arise when the starting point is known, the mediating event is abrupt and obvious, the change from the starting point to the ending point is also abrupt and obvious, and there is no clear, alternative

²Several computer programs, such as NUD*IST and Ethnograph, exist to help analyze qualitative data. Opinion varies on the degree to which these programs are helpful for fairly straightforward case studies. These programs are helpful in finding common themes with a large number (say 10 to 15 or more) of fully transcribed interviews. However, for compiling a list of events and varying perceptions of them, often done in case studies, word processing programs work quite well for keeping track of data. Create a new document for each event and record the differing responses in those documents. The same process works for headings such as Perceptions of Problems, Suggested Changes, and others. Do not invest in purchasing and learning to use a qualitative computer program unless the data set warrants it.

mediating event. Sechrest et al. give as an example a young man who suffered an abrupt blow to the head and experienced a completely changed personality immediately afterward. It is possible in this case to evaluate whether or not the head injury caused the personality change. It is completely plausible that it did. Nothing else occurred, the personality change was abrupt, and the change in the young man was marked. Personalities generally do not change so quickly or dramatically. Part of the certainty in this case stems from the fact that there is a strong theoretical basis for the change, and the change is unlikely to have come about by itself.

Few case studies are this clear, but the same general principles apply to them. If the starting point is clear, the time when the intervention occurred is clear, and if a clear difference exists between the starting point and the ending point, then a defensible case exists that these changes are linked. They occur sequentially, and there is a theoretically strong reason that the change occurred. A theoretically strong reason exists because the program is based on some theory explaining why an intervention should have had the observed effect. The evaluator must be clear about her doubts. In a case study evaluation, these doubts are presented as text in the narrative. If she happens to implement her program in an environment that is changing anyway, she must be honest about the effects of her program relative to other events that were happening concurrently. One way to demonstrate confidence in results is to show that there is no clear, alternative explanation for the observed changes.

For example, an evaluator is trying to counter the effects of tobacco sponsorship at his local rodeo. One measure of success is whether or not there is a decrease in the amount of tobacco gear worn by people over the course of the year. He might even do a gear exchange to help with this measure. Although he could measure change on this dimension by doing a random site survey, this observation might drop into a larger case study in which he interviews rodeo regulars to discover whether or not his presence is known, what links they see between wearing tobacco gear and encouraging tobacco use, and other information of interest. He would also document what he did, and the reaction to his program. At the end of the program, if indeed the wearing of tobacco gear decreased, he has a clear chain of events and a plausible theoretical link between the outcome and the intervention. His final step in interpreting his program as a success is showing that no other plausible mediating event occurred. For example, perhaps in all prior years an industry van came to the rodeo, but it took a pass the year the program was implemented. So while it appeared that the program was victorious in keeping the van away, the van may have been absent for reasons other than the program.

Sources of Error

There are four primary sources of error in case study evaluation: inaccuracies, bias, failure to account for competing explanations, and selection mistakes.

Inaccuracies

Even if every other thing recorded in a case study evaluation is supported and captures reality in a believable way, the evaluator is immediately suspect if the facts that someone can readily check are wrong. For example, is the date right for the passage of a city ordinance? Is the spelling right for the name of the councilor who introduced it? Is the name right for the industry front group that came in to oppose you? If one is sloppy with the obvious facts, then much of the rest of the report is viewed with a great deal of skepticism. It's worth taking time to get it right.

Bias

Bias can enter into case study evaluation in multiple ways. Because one is trying to understand what happened from a variety of perspectives, it's unlikely that one true answer will emerge from the evaluation. Different people see things differently, including the evaluator. These different perspectives can greatly enrich an evaluation, and failure to solicit them can bias it. If there are dissenting views about what happened and the utility of the effort, then one should seek them out.

If an evaluator is also a participant in the activities he is evaluating, his participation is a helpful source of information. He immediately understands the avenues for inquiry, the participants, the setting, and all of the other things that an outsider has to take time to learn. However, if he is this close to the project, then he must be extremely conscious about challenging his pet theories by talking to those who are likely to disagree with him. This challenge is made more difficult by the fact that those who disagree with him may be reluctant to talk to him, knowing he is close to the project

and likely to disagree with them on key points. The failure to pursue other points of view is an important source of error in case study evaluation. The same error, of course, can enter into a case study in which the evaluator is not a program participant, if the evaluator does not seek out divergent points of view.

Competing Explanations

If someone can point to a plausible alternative explanation for the impacts the evaluator attributes to the program AND she did nothing to counter the alternative influence in advance, then she did not complete her evaluation. For example, if she claims that her local program gained voluntary merchant compliance with signage restrictions and resulted in a reduction in signs near schools, then she had better be sure that a state law banning such signage did not pass during the period of her program.

Selection Mistakes

The cases selected for study must, in fact, allow the evaluator to make the points he intends to make. If he wants to talk about factors that lead to the successful implementation of smoke-free bars, then he should present a contrast case, in which implementation was unsuccessful. Otherwise, no evidence exists that the same factors seen in the successful cases are not present in the unsuccessful ones.

Many errors in case study evaluation can be avoided by collecting data from multiple sources, by using several investigators, different methods, different types of data, and/or different sources of data. By taking multiple cuts at the issue, one can avoid some of the problems with data interpretation. Another tool for monitoring case study quality is to ask people who are well informed about the case at hand to review the draft report. These reviewers can be people who were involved in the program, people involved in tobacco control, or anyone else who is positioned to critique the work.

Presentation

A case study evaluation, like all evaluations prepared for the California Tobacco Control Section, should be written up according to the guidelines in *Tell Your Story: Guidelines for Preparing an Evaluation Report*.³ One of the particular difficulties with using case studies to evaluate an intervention is the volume of information collected. By the time observations are completed, interviews conducted, and documents analyzed, it may be hard to see how to write a concise summary.

At this stage the evaluator must sometimes become a brutal editor of his own work. What are, in fact, the key findings? What does the reader need to know? However, the evaluator, as editor, can only go so far. Whatever findings he includes must be documented. He must lay out for the reader the steps he took in doing his work, and the information collected at the various steps, in sufficient detail that the reader will know he covered the essential bases, and that his findings, in fact, are supported by his data. This step is known as creating a transparent path of inference, and its basic function is to let the reader know how the evaluator reached the conclusions he did. The key here is not volume so much as presenting a plausible story that is adequately documented.

Conclusion

A case study is a useful mechanism for evaluating programs, if it is used appropriately and if it is done well. A case study is particularly useful for evaluating unique programs, programs with unique outcomes, and programs carried out in turbulent or unpredictable environments. It is well suited to programs of which at least part of the object is to learn from the process. As with any evaluation, the keys to effective use of a case study are to design the data collection effort with care, and to create a clear record of the methods used to perform the data collection and analysis. Also, as is true of any evaluation, the data must support the findings.

³See footnote 1, page 1

Appendix A

Source Materials for Preparing Case Studies

Miles, M. B. and Huberman, M. A. (1984). Qualitative Data Analysis. Newbury Park: Sage Publications. A very detailed manual of how to analyze data. Helpful for large studies. Probably more than most smaller evaluation studies need.

Merriam, S. (1998). Qualitative Research and Case Study Applications in Education. San Francisco: Jossey Bass. A readable and helpful overview of how to do case studies. Good interviewing section. Although the word, "Education," is in the title, this book has more general application.

Patton, M. Q. (1987). How to Use Qualitative Methods in Evaluation. Thousand Oaks, CA: Sage. A short book on qualitative methods in evaluation, including case studies. Good at explaining concepts.

Patton, M. Q. (1990). Qualitative Evaluation and Research Methods (2nd ed.). Newbury Park, CA: Sage. Patton discusses all form of qualitative evaluation, which he defines as "any effort to increase human effectiveness through systemic data-bound inquiry."

Sechrest, L., Stewart, M., Stickle, T. R., and Sidani, S. (n.d.). Effective and Persuasive Case Studies. Cambridge, MA: Human Services Research Institute. A useful "toolkit" produced by the Evaluation Center. Basic information.

Stake, R. A. (1995). The Art of Case Study Research. Thousand Oaks: Sage Publications. A straightforward look at doing case studies, using as an illustration a case study he did to evaluate school reform in Chicago. This illustration is a good example of what case method can do in evaluation. Stake is not an advocate of audio taping. Good section on identifying issues for the case study.

Weiss, R. S. (1994). Learning From Strangers: The Art and Method of Qualitative Interview Studies. New York: Free Press. A detailed look at interviewing, including examples of transcripts that he critiques. May have more detail than most evaluators need but the examples are interesting to review.

Yin, R. (1994). Case Study Research: Design and Methods. Thousand Oaks, CA: Sage Publications. Yin discusses case study as a research method rather than as an evaluation tool. A good overall resource, however. Yin is explicit about the utility of having quantitative information in a case study.

Using Case Studies to do Program Evaluation

Edith D. Balbach, Tufts University

March 1999

Copyright © 1999 California Department of Health Services

Developed by the Stanford Center for Research in Disease Prevention, Stanford University School of Medicine, in collaboration and with funding from the California Department of Health Services, Tobacco Control Section, under contract number 97-10546. Funding is provided by the passage of Proposition 99, the 1988 Tobacco Tax Initiative.

Special thanks to Carol Chetkovich, Beth Howard-Pitney, and Sue Roberts for their reviews of early drafts of this guide. Thanks to April Roeseler, Carol Russell, and the members of the Local Programs Evaluation Work Group for their input in the penultimate draft.

Design and Production: Stanford Design Group (SDG 945448)