EVALUATION

METHODS FOR STUDYING PROGRAMS AND POLICIES

SECOND EDITION

Carol H. Weiss
Harvard University

P 271 - 293

Prentice Hall, Upper Saddle River, New Jersey 07458

12

ANALYZING AND INTERPRETING THE DATA

> Unobstructed access to facts can produce unlimited good only if it is matched by the desire and ability to find out what they mean and where they lead. Facts are terrible things if left sprawling and unattended.
>
> —Norman Cousins (1981)[1]

> [E]ventually, of course, our knowledge depends upon the living relationship between what we see going on and ourselves. If exposure [to knowledge] is essential, still more is the reflection. Insight doesn't happen often on the click of the moment, like a lucky snapshot, but comes in its own time and more slowly and from nowhere but within.
>
> —Eudora Welty (1971)[2]

Introduction

The aim of analysis is to convert a mass of raw data into a coherent account. Whether the data are quantitative or qualitative, the task is to sort, arrange, and process them and make sense of their configuration. The intent is to produce a reading that accurately represents the raw data and blends them into a meaningful account of events.

This book is no place to learn the techniques of data analysis in either tradition. Many books exist to induct novices into the inner sanctums of analysis and to update "adepts" on the latest analytic developments. The quantitative tradition has 100 years of statistical progress to draw upon. Textbooks are available to explain whether and when to use analysis of variance, analysis of covariance, regression analysis, multilevel analysis, survival analysis, event history analysis, structural equation modeling,

[1]From Norman Cousins, *Human Options: An Autobiographical Notebook*. Copyright © 1981 by Norman Cousins. Reprinted by permission of W. W. Norton & Company, Inc.

[2]From Eudora Welty, *One Time, One Place* (New York: Random House, 1971). Reprinted by permission of Random House, Inc.

and other multi-syllabic analytic techniques. Books on data analysis can be supplemented by excellent journals, university courses, consultants, and conferences.

The qualitative approach, where texts used to concentrate almost solely on field methods for collecting data, has recently seen an explosion of books on analyzing the data collected (e.g., Maxwell, 1996; Miles & Huberman, 1994; Silverman, 1993; Strauss, 1987; Wolcott, 1994). Here, too, there are courses, journals, conferences, and workshops. Much specific advice is available, but the instruction is not (and probably never will be) as detailed or explicit as for quantitative analysis. Despite the plethora of resources, I want this book, unlike other books on evaluation, to say something about data analysis. It is too important a subject to go unremarked.

Analysis grows out of all the preceding phases of the evaluation. It is heavily dependent on the central questions that the study has chosen to address, the research design, and the nature of the measures and data collected. At the point of analysis, the evaluator brings another set of skills to bear in deriving meaning from the design and the data in order to answer the central questions.

In this chapter, I first set out the central tasks of analysis in evaluation. Next I outline a set of analytic strategies common to both qualitative and quantitative analysis, although I go on to point out key differences in the two approaches. Finally, I consider the use of information beyond that collected in the study.

This is a short chapter. It does not try to tell the reader *how* to analyze research data, how to use computer programs or perform statistical calculations, or how to make sense of sprawling descriptions. Libraries are full of books that do just that. At the end of the chapter, I list texts on analytic methods. What this chapter tries to do is provide the *logic* of data analysis in evaluation. Anyone, with or without training in research techniques, ought to be able to follow the logic of the analytic task.

Analytic Tasks in Evaluation

The basic tasks of evaluation are to provide answers to the questions listed in Figure 12-1. An evaluation study can go about answering these kinds of questions in a limited number of ways. Evaluators will follow strategies such as those listed in Figure 12-1. The list presented follows the sequence more common in quantitative than qualitative work. I discuss the qualitative spin on the analytic task in the next section.

No one study will seek to answer all the questions in Figure 12-1, but the figure sets out the range of questions that evaluative analysis can address.

Question 1 calls for descriptions of the program in operation. This is often called process evaluation. The basic task is description. The description can be either quantitative or qualitative in nature, or a mix of both. The quantitative evaluator will rely on counts of various aspects of the program. The qualitative evaluator will develop narratives. Her emphasis is likely to be on the dynamics of the operation and the environment surrounding the program (such as the structure of the program organization or the neighborhood and community environment). As question 1d suggests, she is also apt to be interested in the recipients' interpretations of the program and their participation in it: What does it mean to them in their own terms?

As question 2 notes, the implicit question in a process evaluation may be

FIGURE 12-1 LOGIC OF ANALYSIS IN EVALUATION

1. What went on in the program over time? Describing.
   a. Actors
   b. Activities and services
   c. Conditions of operation
   d. Participants' interpretation
2. How closely did the program follow its original plan? Comparing.
3. Did recipients improve? Comparing.
   a. Differences from preprogram to postprogram
   b. (If data were collected at several time periods) Rate of change.
   c. What did the improvement (or lack of improvement) mean to the recipients?
4. Did recipients do better than nonrecipients? Comparing.
   a. Checking original conditions for comparability
   b. Differences in the two groups preprogram to postprogram
   c. Differences in rates of change
5. Is observed change due to the program? Ruling out rival explanations.
6. What was the worth of the relative improvement of recipients? Cost-benefit or cost-effectiveness analysis.
7. What characteristics are associated with success? Disaggregating.
   a. Characteristics of recipients associated with success
   b. Types of services associated with success
   c. Surrounding conditions associated with success
8. What combinations of actors, services, and conditions are associated with success and failure? Profiling.
9. Through what processes did change take place over time? Modeling.
   a. Comparing events to assumptions of program theory
   b. Modifying program theory to take account of findings
10. What unexpected events and outcomes were observed? Locating unanticipated effects.
11. What are the limits to the findings? To what populations, places, and conditions do conclusions not necessarily apply? Examining deviant cases.
12. What are the implications of these findings? What do they mean in practical terms? Interpreting.
13. What recommendations do the findings imply for modifications in program and policy? Fashioning recommendations.
14. What new policies and programmatic efforts to solve social problems do the findings support? Policy analysis.

whether the program did what it was supposed to do. Did it continue to operate as its designers intended, or did it shift over time? The analytic strategy is comparison. The descriptive data about the program (e.g., frequency of home visits, content of curriculum) are compared to the blueprints of program designers. The answers can indicate the ways in which the program diverged and perhaps why it diverged from the original plan.

Question 3 asks whether recipients of the program prospered. Here the analytic mode is comparing the recipients' situation before and after exposure to the program. If data have been collected several times prior to the program, at several times during the program, and/or at multiple times after the conclusion of recipients' exposure, all those data are included in the analysis. With intermediate data points, the evaluator can assess or calculate the rate of change.[3] For example, did recipients improve rapidly during the early part of their encounter and then level off? Did they improve all across the course of the program so that their immediate postprogram status was much better, only to see the improvement fade out in the succeeding months? Such analysis can be performed either quantitatively or qualitatively.

Question 3c asks about the meaning of program participation to its beneficiaries and how they construed their improvement or lack of improvement. For example, a youngster who barely improved his reading score on a standardized test may still be delighted that he can now read a few simple words and write his name. That may mark tremendous achievement to him. This question is the home turf of qualitative methods.

Question 4 asks how recipients compared to nonrecipients. If the groups were randomly assigned, then their status on key variables was probably very similar at the outset. Any difference at the end represents a gain due to the program. If the groups were not selected randomly, it is useful to find out how their initial status compared, particularly on the critical variable that will be used as the outcome indicator (e.g., frequency of illegal drug use, knowledge of geography). Their status on other factors that might be expected to influence outcomes should also be compared (such as the length of time that they have been using drugs, their IQ scores). If the groups differ in significant ways, this differential has to be taken into account in subsequent analyses.

Question 4b calls for comparing the amount of change in the program and nonprogram groups. If the program group went up considerably on the outcome variable while the comparison group did not change at all, the evaluator has a basis for beginning to make judgments about the value of the program. When the comparison group was chosen through randomization, the evidence seems pretty convincing that the program had some degree of success. But for nonrandom comparison groups, many other possible explanations will have to be tested before reaching that conclusion. Even with randomized controls, additional checks may be required. These kinds of concerns will take us to question 5.

Question 4c is the multi-wave analog to question 3b. It asks for comparison of the rates of change of the two groups. Answers will give an indication of the relative trajectories.

---

[3]John Willett points out that the term *rate of change* seems to imply linear change, whereas much change is nonlinear. I acknowledge the point but retain the term because the alternative—*growth trajectory*—is less familiar to nontechnical readers. Keep in mind that change need not be linear over time.

Question 5 asks whether the program was the cause of whatever changes were observed. The analytic strategy is ruling out plausible rival explanations. As Campbell (1966) has written:

> It is the absence of plausible rival hypotheses that establishes one theory as "correct." In cases where there are rival theories, it is the relative over-all goodness of fit that leads one to be preferred to another. (p. 102)

Could something else *plausibly* have led to the situation found at follow-up? If so, then it is worthwhile examining whether such conditions or events were operative. For example, were the program group and the comparison group different at the outset and could that have been responsible for differential success? *How* did the two groups differ, and is it reasonable to think that differences on *those* characteristics led to different outcomes?

The randomized experiment is the design best suited to ruling out competing hypotheses. Even with random assignment, it is possible that slippage occurred. The seemingly superior performance of program recipients might be due to the fact that the poorer risks dropped out of the experimental group and better risks dropped out of the control group. That would suggest that had it not been for the differential attrition, the program group would not look so good. The basic analytic task is to consider what other influences might have caused the pattern of data observed and then see whether those possibilities can be rejected.

For comparison groups constructed without randomization, many more checks will need to be made. As noted in question 4, the groups might have been dissimilar to start with on the key criterion variables and on variables that affected their ability to do well. If they were located in different communities, they might have been exposed to different influences, which affected their outcomes. In cases such as these, analysis cannot totally compensate for the original inequalities. However, additional analyses can narrow down the list of possible contaminants and help the evaluator reach an estimate of program effects—perhaps not a point estimate but an estimate of the *range* of effects that are probably attributable to the program. The evaluator needs to think hard about all the ways in which the program and nonprogram groups differed not only at the beginning of the program but all through its course and into its aftermath. Then she has to find data that allow her to test the extent to which each factor was implicated in the outcomes.

I remember reading some time ago about an airline crash that occurred because the runway was too short and the plane could not stop before crashing into the barrier. It turned out that the runway had been built to the appropriate length, but over the ensuing months little things happened that chipped away at its size. A new hangar was built that needed extra room. A small part of the runway was diverted to accommodate parked vehicles. The paving at one end was cracking and was removed until it could be replaced. Each of these decrements was small and by itself did not compromise safety. Cumulatively, however, they led to disaster.

That seems to me a good analogy to analysis of evaluation data, especially in cases without randomized assignment. The evaluator has to find each of the small discrepancies that might vitiate the conclusions about the effects of the program. No one of the small problems is important enough to cancel out the first-order effects,

but cumulatively they may undermine the evaluation conclusions—unless taken into account.

Question 6 is the cost-benefit question. It asks for a comparison of the costs of the program with the benefits that accrue to participants and to the larger society. Costs of the program include not only out-of-pocket expenditures for running the program but also opportunity costs—that is, benefits that were forgone because of the program. For example, trainees in a job training program could have been working and earning wages. Because of their enrollment in the program, they lost those wages. That is a program cost.

Program benefits were calculated in answering questions 4 and 5. The analysis there indicated how much and what kind of improvement was the result of the program. For example, an emergency hotline to the police might have increased the number of crimes for which arrests were rapidly made and reduced residents' fears about their safety. The costs of the program can be fairly readily calculated; costs include the personnel staffing the hotline, the phone costs, and the diversion of officers from employment on other police work. What about the benefits? How can the analyst quantify the worth of higher arrest rates and reduced fear? Data are available to estimate the cost of police investigation of a crime, so a decrease in the number and length of those investigations can be costed out. If early arrests are more likely to lead to convictions, then the analyst has to seek data on the benefit to society of placing a criminal behind bars for a given period of time. (The additional costs of prosecuting the cases and of incarceration need to be deducted.) The worth of a decrease in residents' fears is harder to quantify. Analysts have developed methods to quantify the value of such intangible benefits, although as we saw in Chapter 10, much depends on the analyst's judgment.

Question 7 asks which characteristics of the program or the people in it are associated with better or poorer performance. The analytic strategy is disaggregation. The evaluator looks at the data to see if there are apparent relationships between some characteristics of the program and the outcomes. Question 7a asks about participant characteristics (a program input). Did men or women do better? Question 7b asks about program components (program process). Did prison inmates who attended group counseling do better than those who received individual counseling? Question 7c asks about conditions in the environment. Were welfare mothers in low-unemployment states more successful at getting jobs than those in high-unemployment states? The evaluator can look at one variable at a time to see whether any of them proves to be significantly related to outcomes. With regression analysis, she can find out which of a set of variables are associated with good outcomes, when other variables are held constant.

Qualitative evaluators can also examine the relationship between characteristics of clients and the services they received and the characteristics of their progress in the program. Miles and Huberman (1994), for example, provide guidance for constructing matrices that reveal this kind of relationship.

Question 8 asks whether there are clusters of services, conditions, staff, and clients that tend to go along with better outcomes. This question goes beyond the investigation of single variables to examine the effect of combinations of variables. With the use of interaction terms in regression analysis, the analyst can discover the

extent to which the effect of one variable on outcomes is increased or decreased depending on the levels of another variable. For example, single-variable analysis may show that better outcomes are obtained when patients receive service from medical specialists and also when they stay in the hospital for shorter stays. If an interaction term is entered into the regression equation, the results may show that disproportionally better results are obtained when service from medical specialists and short stays are combined. Other statistical techniques allow for investigation of complex relationships in data sets of different types. Similar analyses can be done by qualitative evaluators who systematically factor and cluster the narrative data they have. Ragin (1994) presents a procedure to facilitate such analysis.

Question 9 inquires about the processes through which change takes place. This has traditionally been the domain of qualitative analysis, notably the case study. However, with the development of new methods of structural modeling and growth curve analysis, quantitative methods are catching up to the demands of the task. The evaluator who isn't familiar with these methods can consult a good applied statistician.

One relatively simple way of analyzing the processes of change is though comparison of observed events with the ways in which program actors expected change to occur—that is, comparing events with program theory. The evaluator has laid out expectations for the microsteps by which change will take shape, and she examines evolving events in relationship to that theoretical model. This gives some suggestion about whether the program works in the way that program people anticipated.

If the evaluation data show that most of the assumptions of the program theory were met (e.g., police foot patrol made the police presence more visible to gang members, gang members became more wary of apprehension and reduced the frequency of illegal activities, incidence of crime by juveniles went down), the analysis of why the program worked is fairly well specified. The study may also shed light on other mechanisms at work. For example, another component of program theory may have anticipated that police foot patrol would encourage adults in the neighborhood to take stronger protective action, such as auxiliary neighborhood policing, which would make gang members reduce their criminal activity. Did this chain of events take place? *Why* questions may also be addressed. For example, did gang delinquency simply shift to other neighborhoods that were not being patrolled? Of course, one study is hardly likely to find answers to all the questions that can be asked. But to the extent that data are available to *explain* program outcomes, they can be exploited—to give at least provisional guesses, or new hypotheses, about what processes the program sets in motion.

When the program theory with which the evaluation began proves to be a poor rendition of events, the evaluator has more work to do. Why didn't nutrition supplements to pregnant women improve the birth weight of their newborns? The theory was perfectly straightforward. Nutrition supplements would improve the women's health and nutritional status, and their babies would be born at normal weights. If the data show that this outcome was not reached for most women in the program, do data give clues to the reason? Maybe not. Then further inquiry is required. In one study further investigation showed the women did not take the supplements them-

selves but instead added them to the family's meals—put them in the common pot. Therefore, the supplements had little effect on the woman's health or the birth weight of the child. Unless the quantitative evaluator had foreseen some such eventuality and developed measures to capture such behavior, she might not know why things came out as they did. The qualitative evaluator would have a better chance of finding out, but even she could miss the actions of women back in the privacy of their own kitchens. It might take a follow-on study to track down the mechanisms that led to observed effects.

In using program theory as a guide to analysis, a possible limitation is that the data are not thoroughly accommodating. They may not line up in clear-cut patterns that shout yes, this is the way it is, or no, that's all wrong. The associations are usually going to be partial, and the evaluator will have to exercise careful judgment in figuring out the extent to which they support the theory. One concern that she will have to attend to is whether a plausible theory that has not been tested would fit the data even better. If time allows, she will often want to test several alternative theories to see whether the data fit one of these theories more snugly than the one she started with.

While I have suggested that basing evaluation on programmatic theories of change spins off multiple advantages, it is not a panacea. When the data spraddle in indeterminate ways, they do not allow for clear-cut testing of program theories. The evaluator has to exercise judgment in interpreting how well the program adheres to the posited theories. Statistical tests alone are not going to resolve the issue. Judgment is an indispensable component in the analysis.

Program theory has many benefits for the evaluation enterprise. Even when it does not lead to *crisp* conclusions about the processes and mechanisms of change, it will provide more information and more useful kinds of knowledge than are currently available. Further, its widespread use will encourage analysts to develop and apply new statistical methods to the analysis of the fit between theory and events. Qualitative analysts can hone their own brand of analytic skills on the exercise.

The tenth question focuses on unanticipated (and usually undesired) effects of the program. This question actually has more to do with data collection than analysis. For qualitative evaluators, it suggests that they remain alert to possible undesired side effects throughout the fieldwork. For quantitative evaluators, it suggests that they devise measures of possible unwanted effects early in the game and collect the requisite data. When time comes for analysis, the data of either stripe are analyzed according to the same procedures used for more desirable outcomes.

Question 11 looks outward to the generalizability of study findings. It asks what the limits of the data are, and an analytic strategy for addressing the question is the examination of deviant cases. Let's take an example. An evaluation of a home-care program for the aged finds that most of the clients were maintained in their own homes and were happier there; although their physical condition continued to deteriorate, the deterioration was not as profound as in aged patients in a nursing home. For this program, the results were positive.

The next question is how generalizable these results are to other programs. There may be several features of the study that suggest caution. For example, in one

nursing home evaluation, the study population was all under the age of 80; it is not obvious that results will be the same with older populations. Any limiting condition such as this needs to be specified in the evaluation report. Second, some of the elderly people in the study did not do well. When the evaluator examines these deviant cases, she finds that they tended to be people with serious chronic health problems like diabetes and emphysema. Based on deviant-case analysis, the evaluator may want to limit her conclusions and not claim across-the-board success.

Question 12 asks about the implications of evaluation findings. Putting together everything the study has revealed, the evaluator has to interpret their meaning. She has to decide which among the many data items are important and worth highlighting, and which findings, even if statistically significant, are less noteworthy. She has to combine individual findings into coherent bundles and in some cases disaggregate findings to show the subcategories of which they were composed. For example, Friedlander and Burtless (1995) examined the impact of welfare-to-work programs on participants' earnings. Increase in total earnings was important, but the analysis was more revealing when it isolated the separate effects of a number of components: shorter initial joblessness, faster job finding, increased proportion of people finding employment, earnings on the job, and duration of employment.

Interpretation is the phase of analysis that makes sense out of the varied strands of data. It turns data into a narrative, a story. It identifies strengths and weaknesses and highlights conditions that are associated with both.

The evaluator inevitably comes to the task of interpretation with certain preconceptions. Sometimes the preconceptions take the shape of support for the program, for the general idea behind the program, or for the program's clientele. The evaluator may believe that early childhood education is a very good thing or that job development for adults with disabilities is vitally needed or that adoption of Total Quality Management will make firms better and more productive environments for their workers. Sometimes her preconceptions are more general, taking the shape of perspectives on life, people, research, intervention, justice, and equality. They come from her life experiences and stock of knowledge. At the interpretation stage, they influence the way she proceeds. In the old unlamented positivist world, it was assumed that the facts should speak for themselves and the investigator should be totally neutral. She should do everything she could to expunge her own knowledge and beliefs from the analytic task, to become a cipher while she ciphered. In these postpositivist days, we recognize the futility of such advice; we can't erase the essence of what we are and know from the research process. But we can seek to be conscious of our biases and of the way they affect interpretation, and we can even state our biases up front in evaluation reports.

Much of the knowledge that will influence interpretation is knowledge of prior evaluations of the program or programs of the same type. The evaluator can show the extent to which the current program has been similarly or divergently implemented and the extent to which findings about outcomes support or disagree with earlier findings. The analysis will also seek to answer questions left unanswered in earlier work—in effect, to fill in missing parts of the program puzzle.

Beyond this type of information, the evaluator will bring to bear knowledge from social science research and theory. Concepts and theoretical generalizations

about, say, ego defenses or organizational gatekeeping, can inspire creative ways of interpreting evaluation data. Nor can (or should) the analyst ignore the whole realm of "ordinary knowledge" (Lindblom & Cohen, 1979) that tends to define what we all believe simply because we live in the United States around the turn of the millennium. As Cronbach (1982) has written:

> Experience used in interpreting a summative finding comes partly from supplementary sources. Some of these sources may speak of the program itself, but observation of other programs, tangentially related social research and theory, and commonsense views (community experience) all may carry weight. A study gains authority insofar as it is translated into a story that fits with other experience. (p. 303)

But it is the analyst's job not to accept unquestioningly the everyday assumptions that are part of our workaday lives. She should try to surface them, turn them around in her mind, look at them from a new angle, probe them, and see if the process generates new insights. For example, Schon and Rein (1994) discuss the tenacity of familiar metaphors in structuring how we see the world. They offer the example of fragmented services, a common diagnosis of the problems involved in delivery of social services. This metaphor limits attention to only a subset of the phenomena involved in the problem. It also implicitly contains the outline of the logical solution. If the problem is fragmentation of services, the obvious remedy is integration, making them whole. A commonsense way of looking at the problem thus forestalls other approaches that might have greater chance of success. The evaluator should not settle for facile explanations even when (especially when) they are part of the current orthodoxy.

A dilemma that evaluators may face when they reach the stage of interpretation is what do when the findings show that the program has had little effect. Although, strictly speaking, this is a political rather than an analytic problem, findings of this sort tend to merge the political with the analytic. Especially when the evaluators are being paid by program people and/or have been working closely with them, there is a strong temptation to soft pedal poor results and emphasize positive findings. If the aim is to encourage program managers and practitioners to listen to results rather than reject them and to modify activities rather become defensive, such an approach is understandable. However, when it does violence to the data, it runs the risk of becoming the whitewash that Suchman warned us against in Chapter 2. It is not obvious that downplaying serious flaws will benefit program people in the long run. More helpful will be sensible and practical advice on how to cope with poor results.

Which brings us to the next phase, task 13: developing recommendations for change in program or policy. Evaluators are not always expected to develop the implications of their results into recommendations for action. Sometimes they are asked to be the purveyors of straight evidence, what Sergeant Friday on the old *Dragnet* program called "just the facts, ma'am," while program and policy people see it as *their* responsibility to review the evidence in terms of changes needed. Program people understand the program in its historical setting, its management, the political climate, and the options available, and they can best consider what adapta-

tions should and can be made. They can use evaluation findings as one guide to change, if change seems called for, but they have much other knowledge that they can bring to bear.

In other cases, however, the evaluator is expected to offer her judgment of the types of action that should be taken. The evaluator has an intimate understanding of the details of the information in the study, and she has worked long and hard to comprehend it in all its complexity. Program and policy people often want the evaluator to take the next step and recommend the kinds of changes they should make. When she is called upon to do so, the evaluator has to take full advantage of all the evidence. For example, if she has identified conditions in the program that are associated with better outcomes, she has leads to the kind of recommendations that are well supported in the data. If she has examined the fit of the data to the program's theory of action, the data provide suggestions on steps to take next. If she has data from observation and informal interviewing, she will have ideas about the kinds of changes that make sense.

However, some evaluations are black box evaluations. They have collected data only on outcomes and do not know what went on inside the program box, treating it as though it were totally opaque. The studies may have been after-only, before-and-after, before-and-after with comparison groups, time series, or before-and-after with randomized controls, but the focus was on outcomes, with little attention to the operation of the program. When outcomes are less than satisfactory, the evaluator has little sense of the conditions that led to the outcomes. She is in a bind when it comes to recommendations. Clearly something is wrong now, but what would make it better? Some evaluators have specialized in a programmatic area for years—say physical rehabilitation, child abuse prevention, or teaching English as a second language. Out of their years of evaluating such programs and their other substantive knowledge in the field, they have developed a stock of knowledge. They have a base on which they can draw to make recommendations.

However, many evaluators are not topic-area specialists. They evaluate programs of many different kinds and thus do not have deep knowledge about any one subject area. If the study doesn't provide information about factors that influence program effectiveness, they have to make do with common sense in developing recommendations for change. They are sometimes driven to recommending the opposite of whatever the unsuccessful program is currently doing—especially if the report is almost due and they have little time left to think about the kinds of recommendations that would be sensible.

Needless to say, this is not good practice. Even in this bind, the evaluator has a responsibility to think about the recommendations she is making with the same critical faculty that she devotes to current practices. Are the recommended changes likely to achieve desired effects? Is there experience elsewhere that suggests that they will be practical, feasible, and effective? What program theory would they represent? Are they consonant with other conditions in the environment, such as level of budget or staff capabilities? To find out if the changes she is thinking of recommending are likely to work, she can turn to social science theory and research, past evaluations of similar programs, and the judgment of experts in the field. Collecting information about the likely effects of recommendations is worthwhile, even if it

involves only consultation with program managers and staff. In effect, she would be doing something approaching an evaluability assessment on the prospective recommendations. If there is enough time, a wise course of action would be to collect some data on the recommended practices in action, but except on the smallest scale, this is not likely to be feasible in a time-bound study.

In all cases, however much data the study provides and however substantively expert the evaluator is, the act of making recommendations requires the evaluator to extrapolate beyond the evidence at hand. She has to think about the future world into which any recommended change will enter. If the federal government is going to give states more authority to set rules, if school districts are going to devolve more authority to schools, if budgets are going to be under continued stress, what do these conditions mean for the program in the future? And what kind of recommendations make sense under the coming conditions? If the evaluator has been foresightful, she will have data on some sites that are similar to the conditions that are likely to prevail, such as a state that has been allowed to allocate federal transportation funds across different modes of transportation or a school district that has decentralized decision making to the schools. She can examine these cases carefully as she develops ideas for recommendations.

No one expects the evaluator to be Nostradamus (and how well did all his predictions work?). But the point is that the evaluator cannot assume that tomorrow's world will be the same as today's. Training programs for displaced workers (displaced by shifts in technology) will be affected by changes in economic activity and the employment rate; court-ordered programs of community service for adjudicated delinquents will be affected by changes in laws, sentencing guidelines, and community sentiment. The possibility—even the probability—of relevant change has to be taken into account in the recommendation-making enterprise.

Evaluators rarely recommend the total eradication of a program. Even if data show that the program is totally ineffective, the usual recourse is to try to patch it up and try again. But policymakers will also welcome insights about alternative strategies, and findings from the evaluation can inform the formulation of new interventions. This is the subject of question 13: When new policies and programs are under consideration, what lessons does this evaluation offer to the venture? Here we are in the domain of policy analysis, but we take a distinctively evaluative slant.

The analysis of future policies and programs usually starts with a problem—such as the inaccessibility of public transportation to handicapped people. The policymaker, bureaucrat, or analyst interested in the problem generates a list of possible solutions, such as wheel-chair-accessible elevators in train stations, kneeling buses, and subsidized vans that pick up handicapped customers on call. The analyst has the task of investigating which, if any, of these proposals would be operationally practical, conceptually logical, and empirically supported by past experience. The contribution of evaluation is to supply empirical evidence about the effectiveness of these practices.

The person undertaking analysis of this sort is likely to be a policy analyst rather than an evaluator. But he needs to know a good deal about what evaluation has to say. In analyzing the likely effectiveness of alternatives, he seeks the sum total of evaluation evidence on each topic. Meta-analysis is of particular use. By

combining results of dozens or hundreds of evaluations of similar programs, meta-analysis contributes state-of-the-art knowledge to the forecast. The synthesis of evidence is applied to the analysis of future policies and this extends the reach of evaluation beyond the individual program to the future of programming generally in an area.

Although question 14 takes us beyond the terrain of the individual evaluator, it has been addressed by evaluators. The General Accounting Office's former Program Evaluation and Methodology Division essentially reinvented aspects of policy analysis on a distinctly evaluative base. It published a paper on what it called prospective evaluation, an attempt to formalize a method for extrapolating from existing studies to future times and places (U.S. General Accounting Office, 1989). Because GAO works for the Congress, it is often confronted with requests to examine proposals and bills that would set up new programs, and the paper illustrates the ways in which the staff goes about synthesizing past evaluation findings to judge the likely efficacy of future proposals. An interesting sidelight of the GAO paper is the emphasis it places on surfacing the "theoretical underpinnings of prospective programs" (GAO, 1989, p. 60, see also pp. 30–33). Explicit program theory as well as evaluative evidence help to determine the potential of a proposal for successful operation.

General Strategies of Analysis

The tasks of analysis of quantitative and qualitative data have much in common, but evaluators will go about the tasks in different ways. There are two basic reasons for the difference in strategy: the nature of the data and the analytic intent. First, quantitative and qualitative analysts have different kinds of data in hand. The quantitative evaluator is working with measures that yield numerical values, such as years of schooling or scores on tests of knowledge. Even those data that started out as narrative responses to interviews or notes from observations have been coded into a set of categories to which numerical values are assigned. The qualitative evaluator, on the other hand, is working with sheaves of field notes—narrative accounts of conversations, interviews, observations, fieldworkers' reactions and tentative hunches, school documents, and so on. Therefore, the task that each confronts is different.

Quantitative studies usually have a large number of cases. The large size of the sample helps to ensure that oddities of individual cases cancel each other out, and the data will show the main effects of the intervention. If the sample has been chosen through random sampling, results will generalize to the population from which the sample was drawn. However, quantitative studies often have a limited number of descriptors about each case. A local evaluation can have data on 300 science teachers, and a national study can have data on 2,000, but both have relatively little information about each teacher—perhaps length of time he has been teaching, grade(s) taught, scores on capsule science tests before and after the staff development project, and perhaps one or two more items. Although there is limited information on each participant, the large number of cases makes statistical analysis appropriate.

A qualitative evaluation of the same project would have data on a small sub-

set of the teachers in the staff development project, maybe 8 to 15. If the evaluation team has several members, they might have studied three or four times that many people, but not so many that they could not remember each individual in his own social and historical context. The study would probably have a large amount of data about each one: notes on conversations over time, observations of the person's participation in the staff development project, transcribed interviews, and so on. The information would cover scores of topics, from each person's background and views on teaching to the physical setting of the staff development project, from the actions of the staff developer to the character of the schools to which the science teachers will return. By narrowing down to a relatively small number of cases, qualitative evaluators are able to bring into focus all the details, idiosyncrasies, and complexities in the individual case. However, it is often not apparent how representative these cases are nor how generalizable they are to all science teachers in the program.

The second and more important reason that quantitative and qualitative evaluators undertake different modes of analysis is that their intentions are different. Qualitative evaluators aim to look at the program holistically, seeing each aspect in the context of the whole situation. They are concerned about the influence of the social context. They try to understand prior history as it influences current events. Their emphasis is dynamic, trying to gain a sense of development, movement, and change. In essence, they aim to provide a video (a few years ago we would have said a moving picture) rather than a series of snapshots at time 1, time 2, and time 3. In addition, they are concerned with the perspectives of those involved with the program and the meaning of the experience to them. Rather than impose the evaluator's assumptions and definitions, they seek to elicit the views of participants.

Quantitative analysts traffic largely in numbers. Much of the work that they have to do at this stage they have already done through the development and/or selection of measures and data collection. They have defined the variables of interest and the ways they are measured, and now is the time to focus on locating significant relationships among them. Through statistical techniques, they identify associations among variables and the likelihood that associations are real and not mere chance fluctuations. Their aim is to model the system of cause and effect.

To put the distinction in a more general way, quantitative evaluators tend to focus on whether and to what extent change in x causes change in y. Qualitative evaluators tend to be concerned with the process that connects x and y.

A third difference can be mentioned, the time ordering of analytic tasks. Quantitative evaluators generally wait until the data are in before they begin analyzing them. They need not wait until *all* the data are collected, but they usually wait until all the data *from one wave* of data are collected, coded, checked, entered into the computer, and cleaned. Qualitative evaluators, on the other hand, generally begin their analysis early in the data collection phase. As they talk to people and watch what goes on, they begin to develop hunches about what is happening and the factors in the situation that are implicated in events. They can use subsequent data collection to test their early hypotheses. Analysis is an ongoing process.

Despite the differences in the two traditions, they have common features. Quantitative analysts don't throw numbers unthinkingly into a computer and

swear by the results that come out the other end. Qualitative analysts don't read the entrails of animals to figure out what the data mean. They share commitments to systematic and honest analysis. They also confront common tasks, as outlined in Figure 12-1.

Furthermore, only a limited number of strategies are available for data reduction and interpretation. Listed below is a set of these basic strategies. Most analysts make use of them, whether their proclivities are qualitative or quantitative.

Basic Analytic Strategies

Describing
Counting
Factoring (i.e., dividing into constituent parts)
Clustering
Comparing
Finding commonalities
Examining deviant cases
Finding covariation
Ruling out rival explanations
Modeling
Telling the story

Describing

All evaluators use description to evoke the nature of the program. They describe the program, its setting, staff, structure, sponsorship, and activities. The qualitative analyst is likely to provide narrative chronicles; the quantitative evaluator will provide summary statistics on a series of measures. But in order to give a fair accounting, they all have to engage in description.

Counting

Counting is a part of description and of most other analytic strategies. Counting helps to show what is typical and what is aberrant, what is part of a cluster and what is unique. Even the hardiest qualitative analyst will talk about "more," and "a few," and in order to be sure that she is right, judicious counting is in order. Counts also help to show the extent to which program participants are representative of the larger population to which results might be relevant.

Factoring

I use factoring in the algebraic sense of breaking down aggregates into constituent parts. Quantitative evaluators have already factored aggregates into components in the process of measurement. They have constructed measures to capture different facets of inputs, process, and outcomes. Qualitative evaluators prefer not to abstract people and events into parts but to deal with them holistically. Their emphasis is on gestalt and the natural context. Nevertheless, the human mind, for all its grandeur, is too limited to hold dozens of thoughts simultaneously. Even devout qualitative eval-

uators break out separate aspects of experience and deal with them in sequence—while trying to keep the larger context in mind.

Coding is an example of factoring. The coding of narrative material into categories is a common practice in evaluation. It splices up a narrative into a set of categories that have common substantive content. The procedure allows the analyst to examine the content and frequency of each category, and the covariation among categories.

Clustering

Clustering is the procedure of putting like things together. It identifies characteristics or processes that seem to group, aggregate, or sidle along together. Factor analysis, principal components analysis, cluster analysis, and multi-dimensional scaling are among the statistical techniques for accomplishing this kind of aggregation. Qualitative evaluators tend to cluster cases through iterative reading, grouping, coding, or data displays (Miles & Huberman, 1994). Glaser and Strauss (1967) suggest a systematic procedure called analytic induction for developing categories and models. (I describe it in the section Examining Deviant Cases.) Ragin (1994) describes an explicit way to array qualitative data along a set of dimensions to figure out which cases belong together, what he calls the comparative method. Many qualitative analysts who deal with a small sample of cases rely on their immersion in the materials and repeated rereading of their data to arrive at appropriate clusters.

A concomitant task is deciding what each cluster is a case of. As the analyst develops clusters, she uses intuitive or explicit concepts to group cases together; each case exemplifies the concept that unites them into a category. But as she goes along, she refines and redefines the concepts. Some cases will not fit her original definition of the concept and the concept must be altered or narrowed down, or the case must be dropped from the category and another category found for it. Some concepts will turn out to be useless for making sense of the data. New classifications will be needed. Conceptual clarification is an ongoing task.

Comparing

Comparison is the heart of the evaluation enterprise. The analyst compares the situation before, during, and after the program. She compares program participants with those who did not receive the program. She compares participants with one another to see whether there is a great deal of variability within the program group. Sometimes there is as much variability among the program participants as there is between them and the members of the comparison or control group. In that case, the difference between program and control groups may not be significant, either statistically or practically. Analysis of variance and regression analysis are statistical techniques for finding this out.

Qualitative evaluators, who usually analyze a small number of cases, can compare individuals directly. When they have comparison groups, they, too, have to guard against the possibility that differences within each group are as large as differences between the groups. Rather than make a series of comparisons on single characteristics, they craft multi-dimensional portraits that show similarities and differences.

Finding Commonalities

A basic strategy of the analytic task is to locate common trends and main effects. What are the common experiences that groups of program participants have? What are the common elements that characterize successful participants? Much of the statistical repertoire is geared toward identifying such similarities.

Qualitative analysts who face the same task are often advised to find patterns in the data. Without more specific guidance, I've always thought that that advice is as useful as Michelangelo's advice on how to carve a marble statue: Chip away all the marble that is not part of the figure. One of the specific ways of finding patterns is to look at the frequency of *clusters* of events, behaviors, processes, structures, or meanings, and find the common elements. Brilliant insights arise from finding commonalities among phenomena that were originally viewed as totally unconnected.

Examining Deviant Cases

In statistical analysis, the deviant case is far from the body of the distribution. Analysts may sometimes drop such cases to concentrate on the central tendency and simplify the story. But at some point it is often worthwhile to take a hard look at them and what makes them so unusual. Qualitative analysts, too, may ignore aberrant cases, but important information may lurk at the extreme ends of a distribution. Qualitative analysts can use systematic techniques to incorporate the information into their analysis. One such technique is analytic induction (Bogden & Biklen, 1992; Glaser & Strauss, 1967) which requires analysts to pay careful attention to evidence that challenges whatever constructs and models they are developing. With analytic induction, they keep modifying their concepts and introducing qualifications and limits to their generalizations until the conclusions cover all the data. Cases that are different from the mainstream, outliers in statistical terms, can also give important clues to unanticipated phenomena and may give rise to new insights into what is happening and why.

Finding Covariation

It is one thing to find out whether the values of one phenomenon are related to the values of another. This is the strategy that I have called comparison. Another question is whether *changes* in a phenomenon are related to changes in other phenomena. Does increasing the number of hours of classroom training for disturbed adolescents over the course of the school year lead to better adjustment? Does improving the social relationships among staff in the program lead to improving relationships among participants? Again, statistics has tried and true procedures for making estimates of covariation. Qualitative evaluation has to handcraft such techniques on a study-by-study basis.

Ruling out Rival Explanations

Both quantitative and qualitative evaluators want to be able to ascribe results to the program intervention. They need to separate those changes that the program brings about from changes caused by extraneous circumstances. Even the so-called cadillac of designs, the randomized experiment, may leave loopholes and uncertainties

about causation. The analyst must rule out the possibility that something other than the program was responsible for observed effects.

To go about this step, the analyst identifies plausible alternative explanations. Were successful outcomes brought about by the new math curriculum or by new computing facilities in the schools or the school restructuring that empowered teachers and raised their morale and enthusiasm? Were unsuccessful outcomes the result of the curriculum or were they due to the poor quality of curriculum implementation, insufficient class time devoted to it, or competing demands on students' attention? Whatever factors might *plausibly* be held responsible for observed results—noncomparability of program and comparison groups, outside events, the pattern of dropouts from the program, inappropriate measures, changes in the conduct of the program over time—may need to be examined.

If the evaluator locates a potential explanation that seems reasonable, she has several options. She can collect available data to see whether the rival explanation can be ruled out; she can collect new data to try to disprove it; she can mount a small substudy to investigate its cogency. If the data do not contradict it, she should acknowledge it as a possible explanation for the results and perhaps tag it for further investigation.

Modeling

Modeling is a procedure for putting all relevant information together to create explanations of important outcomes. Here is where all the prior steps come to fruition. The analyst characterizes the nature of program effectiveness. She identifies the elements that were implicated in program effectiveness and assesses the relative weight of their contribution. She notes which elements worked together synergistically to enhance success and which elements were incompatible with each other. To the extent possible, she explains how the program achieved the observed results. She notes which program components and theories were supported in the analysis and which were not. At the end she has a rounded account of what happened and, at least to the limit of her data, how it happened.

Telling the Story

Come the end of the analytic phase, all evaluators have to communicate their findings to readers. Communication requires putting the data into understandable form. Nobody has a monopoly on techniques for making evaluation results clear, putting them in context, pointing out the limits of their applicability, and interpreting them in light of likely developments in program and policy. Writing is where the rubber hits the road. Good writing is not only a matter of style and grace; it is actually a test of the extent to which the writer fully understands what she is talking about. If she can't communicate it well, perhaps she hasn't fully grasped it herself.

An Example of Program Theory as a Guide to Analysis

Program theory can provide a structure for analysis of evaluation data. The analysis can follow the series of steps that have been posited as the course that leads from the intervention through to the achievement of goals.

Let's take an example. A group at Stanford evaluated an education program for asthma patients to enable them to gain increased control over their symptoms and reduce their suffering from asthma (Wilson et al., 1993). The underlying theory was that education would increase their understanding of their condition and therefore their confidence that they could control it, which in turn would lead to their adherence to a treatment regimen, and thus to better control of symptoms. Because the theory guided the evaluation design, the evaluators had measures on each of the steps posited in the theory. The study included 323 adult patients with moderate to severe asthma. There were two treatment groups, one receiving group education and one receiving individualized education, and a control group.

The evaluators monitored participation and found that 88% of the patients assigned to group education attended all four sessions. Observation of the sessions indicated that "educators adhered closely to the program outlines and script, and used the handouts and homework assignments as intended" (p. 569). Notice the importance of this component of the evaluation. Without the monitoring of content, there would be no way of knowing what actually went on in the group education session—that is, whether the program in action was the same as the program design.

The evaluators compared the treatment and control groups on before-and-after measures to see if there were any changes. If not, the thing to explain would be where the theory stalled out and why the program failed. If there were changes, and there were, then the analysis would concentrate on how much change there was, how the program worked, who benefited the most, and which components of the treatment were most effective.

Following the training, patients who received education showed improvements in their test scores on understanding and knowledge about asthma.[4] They also adhered to the regimen of care, which included both cleaning their living quarters and self-care. They showed significantly greater improvement in their bedroom environment (absence of allergenic furnishings, better dust control and cleaning practices) and much better technique in use of a metered-dose inhaler. However, the program group did not significantly improve in getting rid of pets to whom they were allergic or (for the 10% who smoked) giving up smoking, but the number of patients in these categories was small.

Final outcomes for the group-education group were positive. They were less likely to report being bothered by asthma than they had been originally and less likely than the controls; they had fewer days with symptoms than controls; physicians were more likely to judge them improved; their visits to clinics for acute care declined significantly.

However, the article reports the outcomes for the treatment and control groups as a whole and does not link outcomes to earlier stages in the program, such as attendance at group education sessions or adherence to regimen. It is likely that sample sizes would get too small to find significant differences if the group were partitioned along the lines of the theory. Still, it would be informative to know whether those who did everything right (i.e., attended sessions, cleaned up their home environ-

[4]The data about test scores on knowledge about asthma do not appear in the article mentioned. I talked to Sandra Wilson in 1994 and obtained this information.

ments, and used the inhaler properly) turned out to have better outcomes than others, or whether there were some features that didn't link to outcomes. One statement in the article notes that improvement in inhaler technique was associated with reduction in patients' bother with symptoms of asthma, but improvement in inhaler technique "could not account for all of the observed improvement in symptoms" (p. 575). Thus, most of the how and why questions about the program were answered through statistical analysis because the evaluation included appropriate measures to track the unfolding of outcomes.

One rival hypothesis that the evaluators explored was that physicians might have changed patients' medications over the course of the program, which could account for some of the positive results. However, they found that people in the education and control groups did not differ at the outset or at follow-up in intensity of medication (typically four or five daily drugs), and there were no significant changes in the proportions receiving specific types of medications. So medication was not the explanation for change in condition.

A finding about the superiority of group education over individual education led to further exploration of why the group program was more successful. Evaluators interviewed and learned that patients valued and benefited from the opportunity for peer support and interaction, and educators believed that the group setting encouraged patients to express fears and concerns, which the educator could then address. Also, educators were typically more comfortable in group than individual teaching. So here, the why questions were answered through reports from participants.

This report is a fine example of analysis that follows the logical progression of expected effects. It shows that qualitative evaluation doesn't have a monopoly on insight into the processes of program operation. The quantitative evaluator can do just as well if she has savvy about the program, good program theory, measures of inputs and outcomes, and appropriate measures of intervening processes along the lines of the program theory. And good statistical know-how—and the opportunity to add on further inquiry when results are puzzling.

Books on Analytic Methods

It is not possible to shoehorn a description of analytic methods into this book without terminal oversimplification. Many good books already provide the help in this arena that evaluators can use with profit. Here are a few titles.

For Quantitative Evaluation

Linear Models (Regression/ANOVA)

Afifi, A. A., & V. Clark. (1990). *Computer aided multivariate analysis* (2nd ed.). New York: Chapman and Hall.

Lunnenberg, C. (1994). *Modeling experimental and observational data.* Boston: PWS-Kent.

Neter, J., M. H. Kutner, C. J. Nachtschein, & W. Wasserman. (1996). *Applied linear statistical models* (4th ed.). Homewood, IL: Irwin.

Multi-Level Modeling

Bryk, A., & S. Raudenbush. (1992). *Hierarchical linear modeling*. Beverly Hills, CA: Sage.

Goldstein, H. (1995). *Multilevel statistical models* (2nd ed.). London: Edward Arnold.

Structural Equation Modeling

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Loehlin, J. C. (1992). *Latent variable models: An introduction to factor, path, and structural analysis*. Hillsdale, NJ: Erlbaum.

Event History Analysis

Allison, P. D. (1984). *Event history analysis*. Quantitative Applications in the Social Sciences, no. 46. Beverly Hills, CA: Sage.

Blossfeld, H. P., A. Hamerle, & K. U. Mayer. (1989). *Event history analysis*. Hillsdale, NJ: Erlbaum.

Yamaguchi, K. (1991). *Event history analysis*. Beverly Hills, CA: Sage.

Categorical Data Analysis, Log-linear Modeling, Logistic Regression

Agresti, A. (1990). *Categorical data analysis*. New York: Wiley.

Hosmer, D. W., & S. Lemeshow. (1989). *Applied logistic regression*. New York: Wiley.

Meta-Analysis

Cooper, H., & L. V. Hedges. (1994). *Handbook of meta-analysis*. New York: Russell Sage Foundation.

Cook, T. D., H. Cooper, D. S. Cordray, H. Hartmann, L. V. Hedges, R. J. Light, T. A. Louis, & F. Mosteller. (1992). *Meta-analysis for explanation: A casebook*. New York: Russell Sage Foundation.

For Qualitative Evaluation

Strauss, Anselm L. (1987). *Qualitative analysis for social scientists*. Cambridge: Cambridge University Press.

Silverman, David. (1993). *Interpreting qualitative data: Methods for analyzing talk, text, and interaction*. London: Sage.

Wolcott, Harry F. (1994). *Transforming qualitative data: Description, analysis, and interpretation*. Thousand Oaks, CA: Sage.

Miles, Matthew B., & A. Michael Huberman. (1994). *Qualitative data analysis: A sourcebook of new methods* (2nd ed.). Thousand Oaks, CA: Sage.

Maxwell, Joseph A. (1996). *Qualitative research methods*. Thousand Oaks, CA: Sage.

Ethical Issues

Two new ethical issues can arise somewhere around here. One has to do with ownership of the data. The evaluator often thinks that the raw data and analyses belong to her. The study sponsor may believe that his agency owns the data. The disagreement doesn't become contentious until one or the other wishes to release the data to other investigators for possible reanalysis. If a federal agency is the sponsor of the study, it is almost always willing to release data. In some large studies, the government agency will even fund the preparation of public use disks and documentation to make further analysis of the raw data easy. Other sponsors, such as foundations, national associations, and service agencies, can be less disposed to release. The best treatment for this problem is prevention. In the early phases of the study, perhaps even before the contract is signed, the evaluator and sponsor should draw up guidelines about ownership and release (and publication) of the data.

The other problem is the safeguard of confidentiality of the raw data. Even with all the protections that the evaluator has introduced (removal of names and other identifiers), computerized databases are becoming less secure. New technologies are making it easy to access databases, which can lead to the identification of sites and individuals. When databases reach the World Wide Web, the possibilities for data misuse will multiply. The problems are new, and creative people are working on solutions.

Summary

This chapter has discussed analytic strategies for making sense of data. Not seeking to duplicate textbooks devoted to methods of quantitative and qualitative analysis, it tries to reveal the logic of analysis. It sets out a range of questions that an evaluation may be called upon to address and suggests the basic technique that analysis of each question requires.

The basic questions that can be answered in analysis are: (1) what happened in the program; (2) how faithfully did the program adhere to its original plans; (3) did recipients improve: (4) did recipients fare better than nonrecipients of program service; (5) was observed change due to the program; (6) did benefits outweigh costs; (7) what characteristics of persons, services, and context were associated with program success; (8) what combinations or bundles of characteristics were associated with success; (9) through what mechanisms did success take place; (10) what were the unanticipated effects; (11) what limits are there to the applicability of the findings; (12) what are their implications for future programming; (13) what recommendations can be based on the findings; (14) what new policies and programs do the findings support.

In setting out the logic of analysis, I call attention to the commonalities between quantitative and qualitative approaches to evaluation. Both approaches make use of such basic strategies as describing, comparing, disaggregating, combining and clustering, looking at covariation, examining deviant cases, ruling out rival explanations, modeling, and interpreting and telling the story.

The evaluator's analytic priorities are determined by the purpose of the study. The purpose of the study determines the questions asked, the design used, and the

analysis conducted. Take a project that offers clean needles to drug users in order to prevent their contracting AIDS. If the issue in contention is whether drug users will come to the needle exchange project and accept needles from a city agency, then the analysis concentrates on counting and describing those who come to the project and comparing those who attend with the number and characteristics of drug users in the community (from external data sources). If the burning issue is whether rates of new HIV-positive cases go down, then the analysis compares the incidence of cases before and after the initiation of the project.

The chapter also suggests that the evaluator make use of data from sources outside the project to supplement the data collected explicitly for evaluation. Such data can include data from other evaluations, data from social science research, data available in city records or agency files, and even new investigations to follow up anomalies or puzzles that the evaluation reveals. By the end of the analysis, the evaluator should have answers to the questions with which she began.

The chapter concludes with the example of an analysis of a program to educate asthma patients in techniques of self-care.