Now That We've Got Their Attention: Guidelines for Producing Useful and Used Evaluations

Marjorie R. McRae, Barakat and Chamberlin, Inc.
Thom Henneberger and Phil Hanser, Electric Power Research Institute

DSM program evaluations must evaluate issues worth knowing, at a cost worth expending, generating valid conclusions that are used to improve the program and better meet the utility's and customers' needs. Evaluators frequently describe evaluations as comprising three basic activities: methodology planning, data collection, and analysis. To ensure that the results produced are both useful and used, this paper describes how the evaluator might view these activities as part of a larger evaluation process.

The larger process involves six steps. Steps 1 through 4 of this process comprise setting the scope of the evaluation, Step 5 comprises the "doing," and Step 6 communicates and applies the results.

This framework places considerable emphasis on "scoping" the evaluation before detailed methodology planning is conducted and the evaluation launched. This paper argues that proper scoping is essential if evaluations are to be useful and cost-effective. The practitioner must spend time asking questions like the following: To what problem is evaluation a solution? Who are the stakeholders? What information do they need? What analysis approaches will they find convincing? What is at stake? How much can we afford? The most carefully conducted evaluation is meaningless if it asks the wrong questions or answers them in a way that does not meet users' needs.

Introduction

DSM program evaluations increasingly have the attention of utility management and regulators. Evaluations are achieving recognition as critical tools for determining whether DSM programs are meeting utility goals and serving the interests of customers.

DSM program evaluations can serve many needs, such as assessing attainment of program goals and market projections, increasing program effectiveness and efficiency, and calculating program cost-effectiveness. There are also many potential users of evaluations. Uses internal to the utility include program modification, resource planning, load forecasting, rate design, DSM bidding and performance contracting, and market research (Hirst and Reed 1991). Evaluation users also include regulators and, occasionally, consumers. Each of these groups may have different agendas and information needs that evaluations can potentially meet.

With increasing attention paid to evaluation and its application to many utility activities recognized, it is important that evaluators deliver findings that are useful and used. DSM program evaluators must examine issues worth knowing, at a cost worth expending, generating valid conclusions that are used to improve the program and better meet the utility's needs. "What good is a fine evaluation of an important attempt to solve a trivial problem? What good is a fine evaluation of a puny attempt to solve a serious problem? What good is a fine evaluation of a program that solves a serious problem if the results are not stored and used to ameliorate the problem?" (Shadish et al. 1991)

Practitioners frequently describe evaluations as comprising three basic activities: methodology planning, data collection, and analysis. To ensure that results are produced that are both useful and used, this paper describes how the evaluator might view these activities as part of a larger evaluation process. The larger process involves six steps: (1) Identify evaluation objectives; (2) Determine evaluation approaches that will yield information suitable for decision making; (3) Determine possible analytical methods and data requirements; (4) Estimate the maximum expenditure that will still yield a cost-effective evaluation; (5) Conduct the evaluation (following the three basic activities of detailed planning, data collection, and analysis); and (6) Communicate the results.

Steps 1 through 4 comprise scoping the study, Step 5 comprises "doing," and Step 6 comprises effectively communicating the results. Step 5, "doing," tends to receive the majority of evaluators' efforts. A reviewer of this paper suggested that typically 10% of our time goes toward scoping activities, 80% toward "doing" and 10% toward applying the results. However, a division of evaluation effort that more heavily weights scoping of the evaluation and effectively communicating the results is essential if evaluations are to make a significant contribution to enhancing program performance.

Step 1: Identify Evaluation Objectives

The first step of a program evaluation, identify evaluation objectives, is comprised of six substeps, described below. (Hicks [in Hirst and Reed 1991] identifies comparable activities.)

Describe the program. DSM programs are complex, being comprised of elements such as the characteristic of the DSM service or measure (including any technical assistance or training), the use of trade allies in promoting the program, the promotional method and media used, the target market, and the incentive. Because the specification and execution of each of these program design elements can affect program performance, an evaluation should start with a complete description of the program.

Establish a baseline. Next, the evaluator establishes a baseline to provide a description of conditions prior to the program, against which program achievements can be measured. For example, a baseline for an evaluation of promotional effectiveness could include the proportion of customers aware of the DSM technology prior to the utility promoting it.

Identify stakeholders. The program and evaluation stakeholders are the people who will use, are affected by, or have an interest in, the program or the evaluation results. These people include program and other utility staff, participating and nonparticipating customers, trade allies, and regulators.

Identify program objectives. Sometimes, the evaluator may find that program objectives have been explicitly delineated in the program documentation. Whether or not written objectives are available, the evaluator should talk with program designers and implementors to learn what they see as the program objectives. Often, the evaluator may find that different people or groups in the utility hold different objectives for the program--some of which may be poorly defined or even mutually contradictory--and that the objectives may have changed or been added to over time.

Identify uses for evaluation information and accuracy needs. The evaluator needs to identify how the evaluation findings might be used by different groups and the consequent requirements for precision and accuracy. Some decisions (e.g., whether to modify an approach to involving trade allies) may need only qualitative findings. Others (e.g., whether to establish a full-scale program following the pilot program) might require rather precise estimates of program costs and impacts.

Define measurable effects. Finally, the evaluator must specify how the program impacts or effects will be expressed in measurable terms. This step can pose numerous difficulties. For example, in assessing program operational effectiveness, the evaluator must consider how to measure important but somewhat nonspecific program objectives such as whether implementation was "efficient" or whether customer understanding was "improved."

Possible pitfalls in conducting Step 1, identifying evaluation objectives, include the following: (1) Key evaluation stakeholders may not be identified. The evaluator learns of the importance of their objectives for the study after the evaluation is under way or, even worse, upon presenting the findings. (2) The findings may not support decision making and may not product results that can be acted upon. (3) Program and evaluation stakeholders may fail to agree on criteria for program performance, leading to rancor when the results are presented. (4) Program performance criteria may be vague: ways to measure criteria are unspecified, or the measurement criteria are specified, but the score or finding that will count as meeting performance expectations is left vague.

Step 2: Determine the Evaluation Framework

One of the most common pitfalls that arises when conducting an evaluation is that the evaluator fails to develop an explicit research framework. For example, the evaluator may not selected a research approach that sufficiently explains why the observed results were obtained or that reliably distinguishes program effects from other effects on participant behavior. In this second step of the evaluation, the evaluator considers the various types of evaluation frameworks and selects one or more that is appropriate to the evaluation goals and constraints.

Evaluation Frameworks

Evaluators have six general frameworks available to them for structuring DSM program evaluations, as discussed below. Most of these frameworks can contribute to both impact and process evaluations. Each possible framework encompasses many familiar analysis techniques. An evaluator can choose to use several frameworks, thereby investigating program effects using different modes of learning.

Model-based designs. Evaluators who use statistical modeling or regression analysis in their evaluations are working within a model-based design framework even though they may not be aware of it, and their lack of awareness may jeopardize the validity of their research. For example, evaluators may err by attempting to measure a change in behavior (e.g., energy usage) without specifying an underlying model or hypothesis about what causes the observed change in behavior. The model-based design framework is probably the most commonly used evaluation approach and is used most often for load and energy impact evaluations. However, it is also useful in investigating decision making, satisfaction, participation, and free riders. (See, for example, Ozog and Violette 1990.)

Engineering calculations and end-use metering. Engineering calculations and end-use metering estimate electricity use by physically describing, modeling, or metering the electrical loads of the equipment in use in the building under consideration. Engineering calculations go to the source--the equipment--to measure changes in usage patterns. They ensure, for example, that load reduction is not ascribed to a direct load control program that cycles air conditioners off for a shorter period than their duty cycle, or that the savings attributed to each component of a building retrofit program (e.g., roof insulation, window treatments) total less than the heating load. These examples may seem extreme, but statistical methods used without considering the underlying physical relations have been known to lead to such implausible results. However, engineering calculations are only as good as their assumptions about building operations, occupant behavior, equipment commissioning, and maintenance (Nadel and Keating 1991).

Qualitative and quantitative surveys. In impact evaluations, evaluators use surveys to obtain information about such things as participation, persistence, free riders, and free drivers. In process evaluations, evaluators use surveys to determine respondent satisfaction and reaction to program design or activities. Finally, surveys provide data for model-based and engineering studies. Surveys can solicit both qualitative and quantitative information from respondents. Evaluators administer surveys by phone, mail, in person, or through on-site inspection. They summarize survey data using simple descriptive statistics (e.g., reporting the frequency with which each answer was reported or the mean of the sample) using statistical tools to determine the relationships among the variables.

In-depth interviews. Evaluators use in-depth interviews for process evaluations and can even use them to support impact evaluations--for example, to explore issues of DSM measure persistence, free riders, and free drivers. In-depth interviews are particularly valuable for programs in their formative stages to explore responses to new program options and modifications. Guided discussions can investigate such questions as what and how things happen in the program, how the program affects the interviewee, and the ways the interviewee would like to see the program change or thinks it can be improved. Focus groups are a type of in-depth interview: the conversations are unstructured, respondents are free to express their views on a range of issues relating to the program, and the interviewer has the opportunity to continue probing into areas of interest.

Monitoring. Monitoring is the process of tracking and documenting program implementation activities. Although often overlooked, monitoring is the basic building block of program evaluation and should be an ongoing component of program implementation. Without monitoring, the utility cannot determine whether a disappointingly small program impact is the result of ineffectual program design or ineffectual program implementation. Monitoring is particularly important during a program's formative stage for this reason (Rossi and Freeman 1985). Monitoring established programs provides for program accountability and ensures that basic participant data are available in order that a sample might be drawn for more detailed analysis. While perhaps the most common monitoring pitfall is that it is simply not done, some program managers have been known to do it to excess, expending more staff time that in warranted by the value of the information produced.

Observation and case studies. This final evaluation framework examines a small number of cases in great detail. The evaluator observes all aspects of day-to-day program operations, including customer visits. Judgment based on expert opinion, observation, and case studies can be used to answer relatively simple questions with a limited use of evaluation resources or rather subtle, complex questions, such as when extensive observation is paired with surveys and even metering. Willet Kempton (1983, 1984) is a well-known DSM program evaluator who makes extensive use of the observation and case study framework. The main disadvantage of an observation framework is its relative obscurity. Few evaluators take the time required to observe program operations for a long enough period to really understand what occurs.

Selecting an Evaluation Framework

The evaluator should consider the following issues to select an appropriate framework or frameworks.

Would any of the frameworks be unacceptable or unconvincing to stakeholders as a means of evaluating programs? Many evaluators think that a statistical answer is nearly always the best one. Some people, however, believe in the maxim that there are three types of information: "lies, damn lies, and statistics."

Are there precision or accuracy requirements for the data? If so, these requirements necessitate that the evaluator select analysis techniques from among the first three frameworks. Not all analysis techniques within a given framework satisfy the accuracy requirements or even produce accurate statistics. For example, neither a before-and-after bill comparison nor a comparison of the rated efficiencies of equipment (both engineering and metering approaches) can meet stringent accuracy requirements.

Does the evaluation need to answer "why"--that is, why did the observed effect occur, why is satisfaction at the observed level, why were we unsuccessful (or successful)? The evaluator may think that he or she really understands how a program operates as a result of extensive interviews, only to have the program manager ask, "But what about the numbers?" Alternatively, the evaluator may present participation results and be asked, "Can we do it again next year?" If the evaluation needs to answer "why" questions, does it seek a causal explanation or a qualitative understanding?

What is the stage of program development? The various evaluation frameworks are differently suited for programs in different stages of development, as illustrated by Table 1. For example, frameworks requiring more time or cost to implement are less appropriate to programs undergoing rapid change.

What frameworks fit within resource or other constraints? This question can only be answered in general terms, because within a given framework the analytical techniques that an evaluator might employ vary in resource intensity. However, in some circumstances the evaluator might be able to rule out a model-based design because there are few or no observations of customers before participation. In other circumstances, a survey of participants might be out of the question if the participants had just been surveyed for another project.

Table 1. Application of Evaluation Frameworks During Program Life Cycle

	Formative	Early Maturity	Full Maturity
Model-based	3	1	1
Engineering	2	1	2
Survey 1	1	1	
Interviews	1	1	2
Monitoring	1	1	1
Observation	2	3	3

Key: 1 = frequent usage, 2 = moderate usage, 3 = occasional usage.

No single evaluation method is good for all tasks. No study is ever flawless. Evaluators face constraints on the time, resources, and skills available to them for conducting evaluations. The needs of their audience and the extent to which other parties--other utility staff, customers, and trade allies--are willing to participate in the evaluation also impose constraints on evaluators. The point of this step is for the evaluator to consider what needs to be known, what it means to know something in the context of the evaluation, and how each of the frameworks would satisfy or fail to satisfy the evaluation needs. An evaluator must choose among the frameworks with a realistic understanding of the losses and gains to be had from the different approaches.

Pitfalls commonly encountered in this step include the following: (1) The evaluator may not consider the perspectives of the different stakeholders on what it means to "know" something about the program. (2) The evaluator may select the research approach without considering the alternative evaluation frameworks and their advantages and disadvantages in solving the problem at hand. (3) The evaluator may consider statistical techniques without considering a possible underlying model of behavior that would constitute a model-based approach.

Step 3: Determine Possible Analytical Methods and Data Requirements

In this step, the evaluator chooses techniques from the selected framework that might be appropriate to the problem at hand and that reflect a range of resource requirements and information yield. To phrase the task in the vernacular, the evaluator might identify the "Volkswagen," "Chevrolet," and "Cadillac" of analyses. The evaluator then describes the benefits (information yield) and costs (resource requirements) of alternative evaluation methods so that in the next step he or she can select an approach whose benefit to the utility exceeds its cost.

After identifying the data requirements of each technique, the evaluator determines how existing data might be used in the analysis, thereby reducing the cost and time required to complete the evaluation. When the resulting requirements for new data collection are identified and a data collection method proposed, the evaluator characterizes each alternative analytical technique by summarizing its benefits (in terms of type, accuracy, and precision of information obtained) and costs (staff, dollar, and time requirements).

Pitfalls commonly assailing this step include the following: (1) Evaluators may fail to consider using existing data. (2) Existing data may be used without ensuring that they satisfy the evaluation requirements--the converse of the first point. (3) A method may be chosen without carefully considering alternatives, without identifying the type of findings the method produces, and without estimating its approximate resource requirements. (4) The evaluator may incorrectly estimate the availability or cost of acquiring data.

Step 4: Estimate the Value of Information

In Step 4, the evaluator determines the value of the information that the various analytical techniques will produce. An evaluation should provide information that enables the decision maker to know the effects of one or more components of the current program (e.g., its design, implementation, or marketing) and should recommend ways to increase the effectiveness of the program. Armed with this information, the decision maker can compare the consequences of "staying the course" with those of modifying the program in some way. Because a DSM program has value to the utility, any action to "stay the course" or modify the program also has a value. The evaluator needs to understand how utility managers plan to use the evaluation information and the value of the information to the utility. With this knowledge, the evaluator can design a cost-effective evaluation--that is, an evaluation whose cost is less than its value to the utility.

The value of the information may vary according to its accuracy. For example, rough estimates might be useful to one group of evaluation stakeholders but not to another group, which needs more accurate information. More accurate information might be useful to both groups--that is, useful in more decisions--and therefore might have higher value. It follows, therefore, that one analytical technique might produce information of a different value than the information another technique produces. In the words of a prominent evaluator:

> The concept of expected value of information is crucial to evaluation planning. Evaluation information can be very expensive but has the characteristic of diminishing marginal returns. In deciding which information to buy, the evaluation planner must be able to explicitly consider, and trade off, confidence and expected impact within constraints set by the use to which the evaluation will be put. Estimation of information cost and value is the mechanism for tying the evaluation design steps together. (Wholey 1977)

Evaluators can ascertain the value of the information an evaluation will produce by using a method called decision analysis. (See, for example, McRae et al. 1992; Herman 1990; and Pits and McKillip 1984.) The "value of information" approach can be illustrated with an example of the value of information to a decision to offer a compact fluorescent bulb rebate program.

Fictitious Example: The Value of Information

DSM program design staff at a utility are considering whether to offer residential customers compact fluorescent lightbulbs at greatly reduced prices. Several decision makers are opposed to this program, because they believe that customers will install the "funny-looking" lightbulbs in out-of-the-way places like closets, attics, and garages. They also believe that the customers who initially install the bulbs in more frequently used sockets (e.g., living room lamps) might become dissatisfied with them and replace them with traditional bulbs. In short, they believe that even though customers might willingly participate in the program and accept the bulbs, customers will not actually use the bulbs much, which will reduce the program's cost-effectiveness.

The "best case" scenario for the use of the compact fluorescent bulbs entails an average operating time per bulb of six hours a day, 365 days a year. The "most likely case" scenario assumes that customers will use the bulbs four hours a day on average. This scenario is the one that the program designers had in mind. With this assumption, the bulbs have a benefit-cost ratio of 1.2. The "worst case" scenario--the scenario that some decision makers fear--is that customers will only use the bulbs two hours a day on average. Should this "worst case" assumption turn out to be true, the program would not be cost-effective. Table 2 depicts these calculations. The program staff base the net benefits on program costs of $500,000.

The decision makers and designers agree about the likely outcome in the absence of the program. In the "No Program" option, the utility would undertake an intensive media campaign to promote the bulbs but would not offer rebates. Program designers expect that the utility would sell far fewer bulbs, but the bulbs sold would be used for more hours per day on average. The cost of the media campaign to promote the bulbs in the absence of the rebate program would be $150,000. The calculations in Table 3 illustrate how the evaluators assess the "No Program" option.

The net benefits for both decisions (to offer or not offer the program) for each of the usage scenarios are entered into Table 4 to complete the decision analysis. The utility's decision makers would choose to implement the compact fluorescent rebate program because the expected dollar outcome of the program (Box A of Table 4--$100,000) is higher than the expected dollar outcome of not offering the rebates (Box B of Table 4--$70,500).

Table 2. Calculations for a Compact Fluorescent Bulb Program

Cost = $500,000 (media, rebates, and administration)
Expected Number of Bulbs Purchased under Program = 100,000
Benefit-Cost Ratio for "Most Likely" Case Computed to be 1.2
Implied Benefit = benefit-cost ratio × program costs
Implied Benefit for "Most Likely" Case = 1.2 × $500,000 = $600,000

"Most Likely" Case

- Bulb Usage = 4 hours per day on average
- Total Usage of All Bulbs = 400,000 hours per day
- Implied Total Benefit = $600,000
- Net Benefit = $600,000 - $500,000 = $100,000

"Pretty Good" Case

- Bulb Usage = 5 hours per day on average
- Total Usage of All Bulbs = 500,000 hours per day
- Implied Total Benefit = 5/4 × $600,000 = $750,000
- Net Benefit = $750,000 - $500,000 = $250,000

"Best" Case

- Bulb Usage = 6 hours per day on average
- Total Usage of All Bulbs = 600,000 hours per day
- Implied Total Benefit = 6/4 × $600,000 = $900,000
- Net Benefit = $900,000 - $500,000 = $400,000

"Not So Good" Case

- Bulb Usage = 3 hours per day on average
- Total Usage of All Bulbs = 300,000 hours per day
- Implied Total Benefit = 3/4 × $600,000 = $450,000
- Net Benefit = $450,000 - $500,000 = ($50,000)

"Worst" Case

- Bulb Usage = 2 hours per day on average
- Total Usage of All Bulbs = 200,000 hours per day
- Implied Total Benefit = 2/4 × $600,000 = $300,000
- Net Benefit = $300,000 - $500,000 = ($200,000)

Table 3. Calculations for "No Program" Option (Media Only)

Cost = $150,000 (media strategy in absence of rebate)
Expected Number of Bulbs Purchased under "o Program" = 10,000

"Most Likely" Case

- Bulb Usage = 15 hours per day on average
- Total Usage of All Bulbs = 150,000 hours per day
- Implied Total Benefit = 1.5/4 × $600,000 = $225,000
- Net Benefit = $225,000 - $150,000 = $75,000

"Pretty Good" Case

- Bulb Usage = 17.5 hours per day on average
- Total Usage of All Bulbs = 175,000 hours per day
- Implied Total Benefit = 1.75/4 × $600,000 = $262,500
- Net Benefit = $262,500 - $150,000 = $112,500

"Best" Case

- Bulb Usage = 20 hours per day on average
- Total Usage of All Bulbs = 200,000 hours per day
- Implied Total Benefit = 2/4 × $600,000 = $300,000
- Net Benefit = $300,000 - $150,000 = $150,000

"Not So Good" Case

- Bulb Usage = 12 hours per day on average
- Total Usage of All Bulbs = 120,000 hours per day
- Implied Total Benefit = 1.2/4 × $600,000 = $180,000
- Net Benefit = $180,000 - $150,000 = $30,000

"Worst" Case

- Bulb Usage = 8 hours per day on average
- Total Usage of All Bulbs = 80,000 hours per day
- Implied Total Benefit = 0.8/4 × $600,000 = $120,000
- Net Benefit = $120,000 - $150,000 = ($30,000)

These calculations illustrate the decision analysis framework and demonstrate that the evaluators can derive the needed data from a few basic assumptions. Using decision analysis to estimate the value of information need not be time consuming.

Alternatives and Pitfalls

Barnes and McCarthy (1989) describe a different approach to determining how much to spend on an evaluation. Extrapolating from sampling theory, they postulate that evaluation costs increase exponentially as greater accuracy is obtained, while the costs of losses that result from basing decisions on incorrect information increase exponentially as accuracy is foregone. These two relationships define the total evaluation costs (the sum of the two sets of costs for each level of accuracy.) The authors postulate that the optimum evaluation cost is that corresponding to the minimum value of the total cost curve.

Regardless of the approach adopted to estimate the value of information and the upper bound to evaluation expenditures, the practitioner should be aware of possible pitfalls encountered in this step. (1) The evaluation manager might select a research methodology and allocates research funds based on considerations other than the value of information to be obtained. (2) The program staff might not recognize all the costs and benefits associated with the different choices, particularly with a choice not to offer a program. (3) The evaluator might not know whom

Table 4. Value of Information

SCENARIO	Weight	BASIC PROGRAM Program Cost: $500,000 Outcome	Net Benefit	Weight Times Net Benefit	NO PROGRAM Program Cost: $150,000 (Media) Outcome	Net Benefit	Weight Times Net Benefit	WITH PERFECT INFORMATION WEIGHT x MAX. NET BENEFIT
"Best Case" (10%)	0.1	600,000 hours per day	$400,000	$40,000	200,000 hours per day	$150,000	$15,000	$40,000
"Good" (30%)	0.2	500,000 hours per day	$250,000	$50,000	175,000 hours per day	$112,500	$22,500	$50,000
"Most Likely" (50%)	0.4	400,000 hours per day	$100,000	$40,000	150,000 hours per day	$75,000	$30,000	$40,000
"Not So Good" (30%)	0.2	300,000 hours per day	($50,000)	($10,000)	120,000 hours per day	$30,000	$6,000	$6,000
"Worst Case" (10%)	0.1	200,000 hours per day	($200,000)	($20,000)	80,000 hours per day	($30,000)	($3,000)	($3,000)
Expected Dollar Outcome	1.0			A $100,000			B $70,500	C $133,000

Expected value of a decision made with perfect information: $133,000 (Box C)
Expected value of a decision made with no further information: - $100,000 (Max of Boxes A & B)
Expected value of perfect information: $33,000 (Difference)

to ask or whom to believe in identifying the possible outcomes and their associated probabilities. (4) The program staff might miss the opportunity to use the decision analysis information in connection with other pending strategic program decisions.

Step 5: Conduct the Evaluation

This step comprises those activities typically considered to constitute program evaluations--methodology planning, data collection, and analysis. The evaluator begins Step 5 by selecting for the analysis an evaluation technique whose value to the utility exceeds its cost.

Many books and articles have been written on techniques for conducting DSM program evaluations. The field is too vast to summarize in a few paragraphs here. Interested readers should investigate such sources as Spinney et al. (1992), Hirst and Reed (1991), Violette et al. (1991), and proceedings from this biennial conference and the biennial Chicago Evaluation Conference.

Obviously, the pitfalls common to conducting an evaluation depend on the analytical techniques selected. The above sources discuss mistakes practitioners will want to avoid. Pitfalls that to many different analyses include the following: (1) The evaluator might initiate the evaluation without coordinating the timing or data collection needs with other planned evaluations. (2) The evaluation might not prove whether the program caused the observed change. (3) The evaluator might not have explored analytical issues raised during the evaluation because he or she is too attached to the initial plan. (4) The evaluator might have allotted insufficient time for data cleaning and construction of an analysis database.

Step 6: Communicate the Results

Peter Rossi, a noted evaluator, states, "Successful evaluators are those who have made clear to themselves, and to their sponsors and program staffs, how the evaluation is to be used and its level of application" (Rossi and Freeman 1985). Too often evaluators are disappointed to see their reports "sitting on a shelf," the recommendations unimplemented and the basic findings unknown except to those who requested the study.

Most evaluation findings have implications for many utility activities, as discussed in Step 1. Nonetheless, sometimes the greatest evaluation challenges lie in communicating the results in such a way as to stimulate their use. In addition, the evaluator faces difficulties in communicating the significance and implications of results to utility upper management and regulatory staff who may have only a general understanding of evaluation techniques. Such evaluation recipients may distrust evaluation findings, not understand how to apply them, or feel uncomfortable with evaluation findings that include caveats or uncertainty bands around the estimates.

The following approaches can increase the likelihood that evaluation results are used throughout the utility.

During evaluation scoping. Identify the potential users of evaluation results at the outset. Ask staff about how they use information. Study program components over which evaluation users have control. Include plans for utilization and dissemination of results as part of the evaluation design. Allocate a portion of the evaluation budget sufficient to support the dissemination of findings through multiple reports and presentations.

While conducting the analyses. Maintain contact with evaluation users throughout the evaluation. Discuss with decision makers the desirability of investigating unanticipated issues or effects that become apparent during the analysis. Provide interim results.

Describing findings. Translate findings into recommendations for action (e.g., modifying program design, implications for performance contracting). Express the error bounds of the impact estimate in the units of the impact (e.g., in kilowatts). Decision makers who think that an uncertainty of $\pm 30\%$ is unacceptable often think that an error band of 0.3 kW is trivial, even when the point estimate is 1 kW. As Buller and Miller (1992) point out in their discussion of uncertainty in the impacts of aggregate DSM impacts, "long term planning at PG&E can't even see effects smaller than 50 MW."

Disseminating findings. Ensure that results are timely and available when needed. Prepare different reports for different levels of management. Write reports with the audience in mind. Avoid overwhelming decision makers by providing to much detail. Write brief, nontechnical executive summaries that contain action recommendations. Disseminate results through informal meetings and oral briefings. For programs benefitting public institutions (such as schools or hospitals), consider publicizing results through media presentations. Circulate results to other researchers and people interested in the issue.

In addition to the above measures that an evaluator can take, certain conditions also contribute to the acceptance and use of evaluation results within the utility. First, evaluation results are more likely to be accepted if the organization conducting the evaluation has a high degree of credibility. For example, Oak Ridge National Laboratory conducted what became highly publicized evaluations of the weatherization and Hood River programs of the Bonneville Power Administration (BPA).

Second, evaluations of high-profile DSM programs often receive considerable attention. High-profile programs may be characterized as those that have large estimated impacts or large budgets, relate to the current "hot topic," are experimental, or are widely known within the service territory (such as the residential lighting programs of Boston Edison and Central Maine Power that involved Lions Clubs.)

Third, continuity within the utility's evaluation staff can ensure that evaluation results are not forgotten should the issues addressed by the evaluation become "hot" at a later time. For example, the majority of an evaluation of BPA's Sponsored Design Program had been completed for four years, and the report from the last phase of the evaluation had languished on the shelf for one and a half years when, nationwide, interest was piqued in bidding and industrial programs. BPA's program evaluation manager turned to the completed research to provide the utility with insights and findings and even revisited the project data to answer new questions.

Conclusions

As evaluators have long recognized, evaluations provide the information needed to increase the effectiveness and efficiency of DSM. And in fact, evaluations increasingly have the attention of utility management and regulators. However, now that we have their attention, we need to ensure that our evaluations themselves are effective and efficient. Given the many issues that can be studied in our quest to improve program performance--and the necessarily limited amount of resources that can be allocated to evaluation--it is important that we evaluate issues worth knowing, at a cost worth expending, generating valid conclusions that are used to improve the program and better meet the utility's and customers' needs.

This paper shows that undertaking scoping activities prior to conducting the evaluation, and effectively communicating the results after we have conducted the evaluation, are critical for producing evaluations that are both useful and used. Many of us have had experience with evaluations that seem to go nowhere but someone's shelf. We need to take the time to seriously answer such questions as: What is the problem or situation that the evaluation is addressing? Who are the decision makers? What information is necessary to support decision making? What is the best match between evaluation tools and stakeholder needs? How much evaluation is cost-effective? What forums for presenting the results will increase their acceptance and use? This paper offers guidelines for asking and answering these questions. We may find that addressing these issues is as challenging and interesting as "doing" the analyses that, up until now, have occupied most of our time.

Acknowledgements

This paper benefitted substantially from the comments received by reviewers--Elizabeth Hicks, Vince Schueler, and Jane Peters.

References

Barnes, H. H., and P. McCarthy. 1989. *"How Much to Spend on Evaluation: What is the Value of Information?"* Applied Management Sciences, unpublished.

Buller, S., and W. Miller. 1992. "How Should We Treat Factors Contributing to Uncertainty in Measurement and Evaluation of DSM?" *Proceedings from the ACEEE 1992 Summer Study on Energy Efficiency in Buildings*. Draft. American Council for An Energy Efficient Economy, Washington, D.C.

Herman, P. M. 1989. "An Algorithm to Determine the Value of End-Use Load Information." *Proceedings of the Second Annual Conference on End-Use Load Information and Its Role in DSM*. The Fleming Group.

Hirst, E., and J. Reed, eds. 1991. *Handbook of Evaluation of Utility DSM Programs*. ORNL/CON-336. Oak Ridge National Laboratory. Oak Ridge, Tennessee.

Kempton, W. 1983. "Two Theories of Home Heat Control." *Cognitive Science*, 10:75–90.

Kempton, W., C. K. Harris, J. G. Keith, and J. S. Weihl. 1984. "Do Consumers Know 'What Works' in Energy Conservation?" *What Works: Documenting Energy Conservation in Buildings*. J. Harris and C. Blumstein, eds. American Council for an Energy-Efficient Economy, Washington, D.C.

McRae, M., P. M. Herman, J. S. Peters, and A. A. Goett. 1992. *DSM Evaluation: Six Steps for Assessing Programs*. EPRI CU-6999. Electric Power Research Institute, Palo Alto, California.

Nadel, S. M. and K. M. Keating. 1991. "Engineering Estimates vs. Impact Evaluation Results: How Do They Compare and Why?" *Energy Program Evaluation: Uses, Methods, and Results: Proceedings from the Fifth International Conference in Chicago*. CONF-910807, pp. 24-33. U.S. Department of Energy, Evanston, Illinois.

Ozog, M., and D. Violette. 1990. "Using Billing Data to Estimate Energy Savings: Specifications of Energy Savings Models. Self-Selection and Free-Riders." *Proceedings from ACEEE 1990 Summer Study on Energy Efficiency in Buildings*. Volume 6, pp. 131–140. American Council for an Energy Efficient Economy, Washington, D.C.

Pits, G,. and J. McKillip. 1984 *Decision Analysis for Program Evaluators*. Sage, Beverly Hills, California.

Rossi, Peter H., and H. E. Freeman. 1985. *Evaluation: A Systematic Approach*. 3d ed. Sage, Beverly Hills, California.

Shadish, Jr. W., T. D. Cook, and L. C. Leviton. 1991. *Foundations of Program Evaluation*. Sage, Newbury Park, California.

Spinney, P., P. O'Rourke, and J. Peters. 1992. *DSM Process Evaluation: A Guide to Current Practice*. EPRI TR-100394. Electric Power Research Institute, Palo Alto, California.

Violette, D., M. Ozog, M. Keneipp, and F. Stern. 1991. *Impact Evaluation of Demand-Side Management Programs*. Volume 1: *A Guide to Current Practice*. EPRI CU-7179. Electric Power Research Institute. Palo Alto, California.

Wholey, J. S. 1977. "Evaluability Assessment." *Evaluation Research Methods: A Basic Guide*, L. Rutman, ed. Sage, Beverly Hills, California.