Building Reliable DSM Resources with Program Evaluation

Michael Sedmak, Robert Uhlaner, and Bruce Smith, Quantum Consulting Inc.

Interest in demand-side management (DSM) as an important component of utility integrated resource portfolios has been steadily increasing (See Association of DSM Professionals, 1993). The basic premise of this paper is that when DSM program planners, implementors and evaluators, as well as integrated resource planners, work in concert, evaluation actually helps the utility *create usable DSM resources.* In many instances, however, DSM programs are fielded without an understanding of the contribution of program evaluation to the establishment of DSM as a viable utility resource. The paper will show that the value of DSM to a utility is critically dependent on the degree to which integrated resource planners, DSM program designers, implementors, and evaluators interact.

Introduction

As has been discussed by others, the value of DSM resources to utilities *and* their customers relies on evaluating them on a "level playing field" with traditional supply-side and other resource options (See Chaisson and Coakley 1992, Kushler et al. 1992, and Lemaye and Davis 1993). Establishing this playing field requires a number of actions, including: establishment of concise DSM program objectives, inclusion of all resource costs in the IRP comparisons, and recognition of the uncertainty regarding supply-side resource benefits and costs.

Evaluators often face resistance from DSM program designers and implementors, who sometimes feel that spending money on evaluation is both depleting the available pool of DSM resources and is getting in the way of successful program implementation. Among the comments sometimes heard by evaluators during initial contact with DSM program designers and implementors are: (1) "You're stealing from me''–That is, every dollar spent on evaluation detracts from the available pool of DSM resources. Addressing this concern is the main topic of this paper; (2) "Who put you in charge? "-The issue here is that evaluations are sometimes viewed by the people "being evaluated" as management audits. It is up to the evaluators to assure their internal utility clients that evaluations are not designed to judge performance, but instead are undertaken to produce analytic results that can be used to enhance DSM program effectiveness.; (3) "Leave our customers alone!"-The issue here is a very legitimate and important concern about avoiding unnecessary customer contact. The most successful response to this issue is to work with program designers and implementors to develop a cohesive survey research plan, so that the needs of all parties are met as efficiency as possible, and that evaluation-related customer contact is minimized.

In spite of concerns such as these, however, we have found that when evaluators make a concerted effort to work with their internal clients throughout the evaluation process, evaluators can actually help to create an environment in which evaluation results are used to enhance the value of the DSM resource.

In this paper, we illustrate that the "usable" DSM resource is directly related to the accuracy with which current and future program costs and benefits can be measured through monitoring and evaluation (M&E). This approach provides a systematic framework as an alternative to the well-known, but strictly *ad hoc,* "10% Rule" for allocating DSM program budgets to M&E. The paper concentrates on describing how the development of tangible DSM resources can be realized with *systematic program evaluation.* That is, rather than detracting from the amount of DSM resource "produced," evaluation actually helps to create usable DSM resources. The paper begins with a discussion of the role of evaluation in creating a DSM resource from a system planning perspective. The steps that should be undertaken to ensure a successful evaluation are then discussed. The steps include: (1) Specifying the utility's DSM evaluation objective function, (2) Determining the relative importance of the programs to be evaluated, (3) Assessing existing evaluation estimates, (4) Estimating the cost of improving the estimates, (5) Estimating the value of improved estimates, and (6) Determining the "optimal" level of evaluation. From these steps, decision-makers can determine the appropriate evaluation budget.

"Usable" DSM Resources

Investing in evaluation does detract from the amount of DSM resource produced, but only in a world of perfect knowledge. The apparent dilemma facing DSM resource "providers" when initially confronting the need for evaluation is presented in Figure 1. In general terms, every dollar spent on evaluation will detract from the amount of DSM resource potentially available, if we know with certainty exactly what the actual program impacts (the resource) are. This is never possible. In the figure, with no evaluation the "certain" impact is about 900 MW; but the confidence interval around that impact is plus or minus 800 MW. In addition, there is some natural variation in the amount of DSM resource produced. In the figure, this is labeled "actual variability in impact." Just as is the case with supply-side resources, there are imperfections in all DSM technologies that result in somewhat uncertain performance. Although behavioral factors and climatic conditions also affect the performance of supply-side resources, these factors most likely have a greater effect on the performance of demand-side resources, making the underlying uncertainty in DSM resources somewhat greater than that of supply-side resources. The only time at which there is no variation in the DSM (or supply-side) resource is when there is no resource, that is, when the DSM implementation investment is equal to zero. Thus, we believe a principal value of DSM evaluation is in reducing the uncertainty associated with the DSM resource.

In fact, DSM program evaluators are not "stealing" from DSM resource providers, but are instead assisting them, and the utility to create a more viable, more usable utility resource. The level of "usable" DSM resource is much higher under a DSM investment portfolio that includes an appropriate amount of evaluation than it is without evaluation. One of the most important questions evaluation attempts to answer is "What is the DSM resource really worth?" In order to answer this question, the evaluator must understand the method by which the utility in question values its DSM resource. In the current example, it is assumed that the value of the DSM resource is related to the level of confidence at which DSM program demand impacts have been estimated.

Specifying the Evaluation Objective Function

Given knowledge of the utility's evaluation objective function, the relationship between "usable" DSM resources and the level of investment in evaluation becomes the "cornucopia" shown in Figure 1. As illustrated. the accuracy at which DSM program impacts can

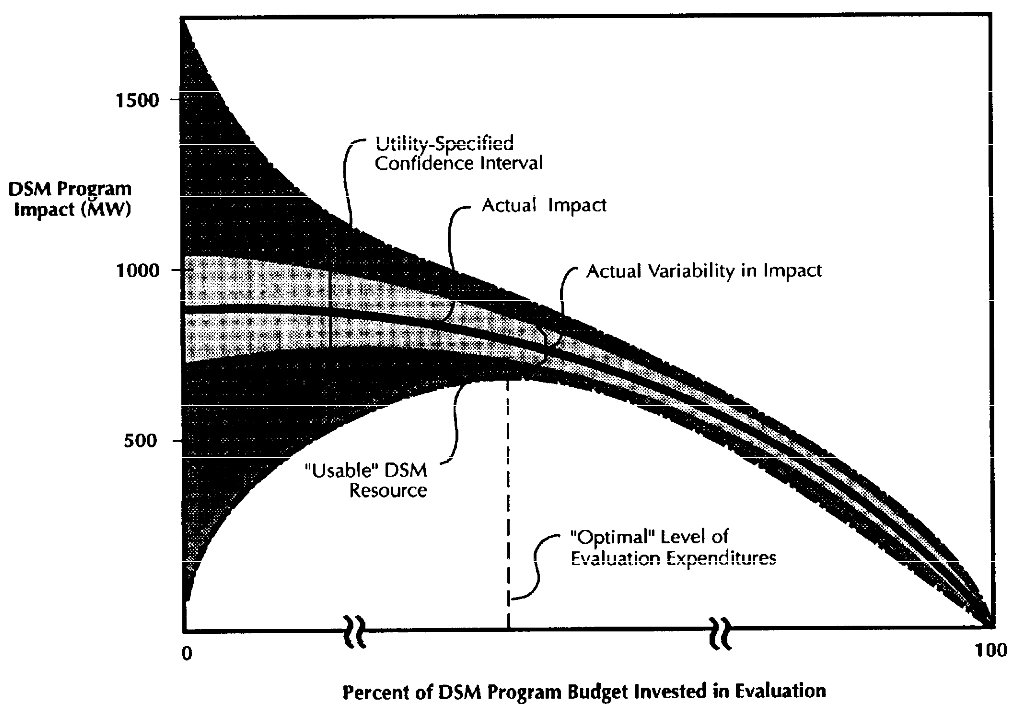


Figure 1. Determining the "Optimal" Level of Evaluation

be measured increases with the level of investment in evaluation. Given a measurable utility objective function, the relationship between the value of the DSM resource and the level of investment in evaluation can be specified, and the "optimal" level of evaluation can be determined. In this illustration it is assumed that the utility has defined a "usable DSM resource" conservatively as the lower bound of the 90 percent confidence interval. This definition was made through discussions with its integrated resource planning (IRP) staff. Of course, other definitions of "usable resource" are possible. Given this assumption and the resources required to estimate demand impacts at different levels of accuracy, the "optimal" level of evaluation can be determined, as shown.

Furthermore, this concept illustrates the importance of requiring certain levels of evaluation performance when inter-utility comparisons are made. In an example where potentially significant utility revenues could be at stake, the Renewable Energy Reserve Credits that are available under the 1990 Clean Air Act Amendments can be calculated using evaluation protocols approved by individual state utility regulatory commissions. Since individual regulatory commissions can have very different evaluation requirements, inter-utility differences in allowed DSM program benefits in part may be due to the different levels of accuracy at which evaluations have been conducted. (Wilems et al. 1993).

In order to determine the "optimal" level of evaluation, the utility's objective function must be understood. Since this function cannot be exactly specified, the most effective way to determine the utility's objective function is to interview the internal users and clients of the evaluation. Senior management and operational staff in the utility departments concerned with DSM should be interviewed including, for example, marketing, system planning, regulatory affairs and field staff. Ideally, the interviewees are initially asked to list the primary uses of evaluation from their perspective. Their open-ended responses are then assimilated into a set of possible objectives, and the interviewees are then asked to rank the importance of each objective. It is up to senior management to weigh the contribution of each participating group to the determination of the final objectives and their priorities.

Another valuable benefit of this approach is that the internal clients of the evaluation become involved from the beginning, thereby providing the users of the evaluation with at least some "ownership" of the process. That is, the success of any evaluation is directly related to the degree to which the utility's decision makers support the evaluation process and its uses. It is important, therefore, for evaluators to be proactive, and to work with DSM program management, designers, implementors and system planners throughout the evaluation to ensure that the evaluation results will be as "used and useful" as possible. As mentioned above, senior management input into the process is very important. The importance of assessing customer satisfaction as an evaluation objective would be much higher if the responses of upper management received higher weights.

One issue that sometimes arises during the course of these interviews is that DSM program priorities, and therefore evaluation priorities, change often. Just as the future value of evaluation resources cannot be measured exactly, nothing can be done about unforeseen changes in priorities. What can be done, however, is to get the clients of the evaluation to specify and prioritize the planned uses of the evaluation at the beginning. If flexibility is desired, the evaluation plan can be specified to provide the required flexibility. It is important, however, for the users of the evaluation results to understand that for any given amount of evaluation resources the flexibility is accomplished at the expense of the accuracy at which specific goals can be met. Furthermore, by getting evaluation users to specify their likely uses of evaluation results at the beginning, the likelihood of substantial change during the course of the evaluation is minimized.

Figure 2 illustrates how key evaluation objectives can be set on a program-by-program basis, as well as by type of evaluation (e.g., impact, market, and process) and over time. For example, the figure shows that the utility's initial Residential A/C Retrofit program net kW/kWh impacts required low precision and level of evaluation effort, but that the fill-scale program evaluation would require high precision and level of effort. Similarly, other types of evaluation objectives and programs are also prioritized.

Assessing the effects of the DSM programs on future market penetration is very important to a utility, for two reasons: (1) analyzing market penetration *over time* is the best way to estimate net impacts; (2) the utility is interested in assessing the efficiency of current program design. That is, penetration under alternative scenarios needs to be assessed.

One of the reasons that assessing the effects of DSM programs on future market penetration is a high-priority evaluation objective is that technology diffusion with and without the DSM program must be estimated to determine net program benefits. Although it is true that DSM efficiency programs heighten awareness of the benefits of energy efficient investments, DSM programs themselves do not "create" resources. Instead, the programs accelerate the adoption of energy efficient measures, as illustrated in Figure 3. Once estimates of penetration with and without the DSM program have been made, estimating the net benefits of the program merely requires calculating the

Key Evaluation Objectives	Residential A/C Retrofit, A/C Load Managment, Commercial Lighting			Custom Rebate, Commercial Audits		
	Evaluation Year			Evaluation Year		
	Initial	Intermediate	Full Scale	Initial	Intermediate	Full Scale
IMPACT						
Measuring Net kW/kWh Impacts	○	◍	●	○	◍	◍
Measuring Gross kW/kWh Impacts	◒	◍	●	◒	◍	◍
Measuring Impacts by Segment	◒	◍	●	◒	◍	◍
Assessing Persistence of Measures	○	◒	●	○	◒	◍
Assessing Free Ridership	○	◍	●	○	◍	●
MARKET						
Assessing Effectiveness of Current Program Design	○	◒	◍	○	◒	◍
Assessing Future Market Potential	○	◒	◍	○	◒	◍
Analyzing Penetration Under Alternative Scenarios	○	◒	◍	○	◒	◍
PROCESS						
Assessing Efficiency of Program Delivery	○	◒	◍	○	◒	◍
Assessing Customer Satisfaction	◒	◍	●	◒	◍	●
Assessing DSM Staff/Trade Ally Performance						

Key	
Precision/Level of Effort Expended	
●	High
◍	Medium-High
◒	Medium
○	Low

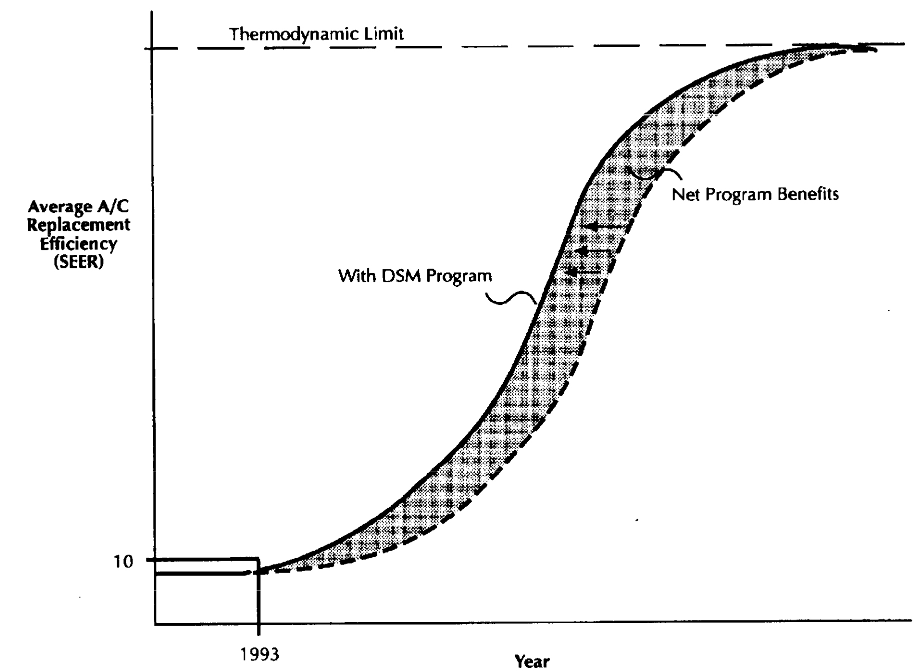
Figure 2. Key Evaluation Objectives by Program and Evaluation Type

present value of the area between the two penetration curves, the present value of the shaded area in Figure 3.

Although concepts such as free riders and free drivers are very useful in explaining the possible effects of DSM programs on the penetration of DSM measures, it is not necessary to attempt to estimate the "level" of free ridership or free drivership when estimating net program impacts. As discussed above, it is possible to estimate market penetration with and without the DSM program. What we cannot do very easily is estimate the free driver effect—the effect the program has in moving the market. Therefore, since we really cannot determine the level of free drivers, it is not possible to (easily) estimate the level of free ridership, either. Luckily, this apparent problem really isn't all that important, as what matters in the evaluation is determining net program impacts. All that is needed to determine net impacts are estimates of penetration with and without the program. Just like it is not necessary to measure rebound to estimate gross impacts, it is not necessary to measure the *amount* of free ridership or free drivership (or the number of free breakfasts, lunches and/or dinners, for that matter) to determine net impacts.

Determining Program Importance

After the objectives have been specified, the relative importance of the candidate programs to be evaluated can be determined. In the example used in this paper, estimating gross demand impacts and assessing the effects of the programs on future market penetration are the highest priority objectives, the program rankings presented in Figure 4 are based on these objectives. The vertical axis



The vertical axis in this figure is labeled "Average A/C Replacement Efficiency (SEER)" for illustrative purposes. When net program impacts are actually calculated, the present value of the difference in energy usage and/or demand with and without the program is calculated.

Figure 3. Effect of DSM Program on Diffusion of Energy-Efficient Technologies

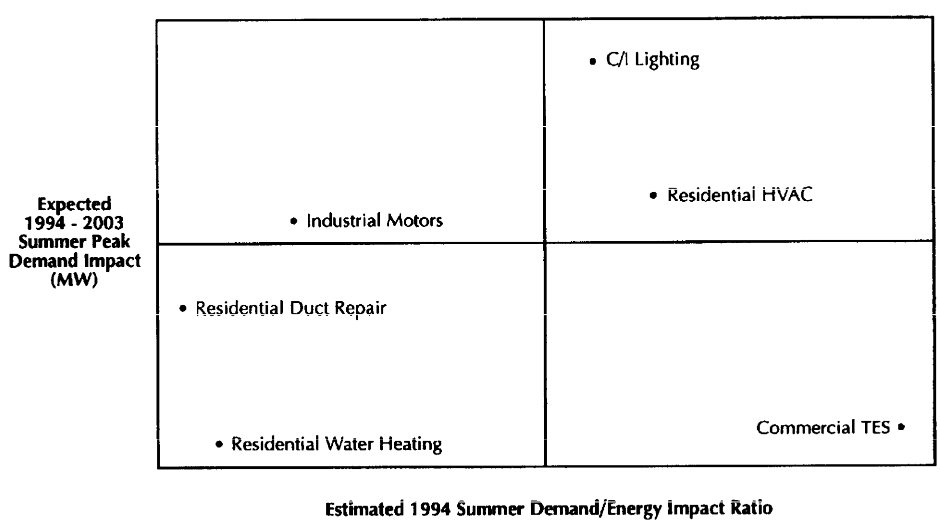


Figure 4. Relative Importance of Programs - Demand Impact Potential vs. Impact Load Factors

in Figure 4 represents expected market penetration over the next decade times the expected per participant gross demand impact. The horizontal axis is (l/Summer Load/ Impact Factor), and indicates the "peakedness" of the program impacts. That is, programs with higher Summer Demand/Energy Impact Ratios are higher-priority candidates for program-specific load research than are programs with lower ratios.

It would be possible to perform these rankings on any of the objectives. In fact, a weighted average of the rankings of each program in accordance with each of the objectives would be ideal, and one that we have used. Basing evaluation priorities primarily on lower-ranked (but less resource intensive) objectives such as measuring gross kWh impacts is, however, inappropriate. In fact, many evaluations stress realized gross kWh impacts, as estimation of these impacts (at least for programs with relatively high levels of participation in homogeneous segments) can be performed with well-known econometric techniques, and ignore other, sometimes more important, objectives, such as net impact calculations, shown to be extremely important for the utility in question.

Two more caveats about the use of this method are worth mentioning. It is important to remember that the methods presented here are really decision-making tools. As stressed earlier in this paper, all final decisions should be made by senior management. The rankings presented are based on currently available program design and/or previous evaluation estimates. As evaluations proceed and more information is available, the rankings themselves may change. For example, a Residential Duct Repair Program to be evaluated received a relatively low priority, as is has been found that per participant demand impacts for duct repair programs are very low at time of system peak. Prior to this finding, however, the Duct Repair Program would have received a much higher ranking.

Assess Current Impacts

Once the evaluation objectives have been specified and prioritized and the programs to be evaluated have been prioritized, an assessment of the current estimates of program impacts can be conducted. That is, the evaluators must figure out where the starting line is before they start running. The results of such an effort are presented in Figure 5. Note that more is known about gross energy impacts and customer satisfaction that is known about the relatively higher-ranked objectives, estimating gross demand impacts and assessing the effects of the programs on future market penetration. This is not surprising as it is possible that the interviewees included some assessment of the current state of knowledge about each objective in their ranking. Regardless, it is apparent from Figure 5 that enhancing estimates of gross demand impacts and future market penetration will have the highest value to the utility.

Cost of Improving Impact Estimates

The costs of obtaining different levels of accuracy in the analytic outputs required to meet each objective are then specified. We have developed a series of methods for estimating the costs of obtaining different levels of accuracy for selected evaluation objectives. An example for load data collection and analysis is presented here. Estimating a cost function for the measurement of each major evaluation objective is an essential component of an actionable evaluation plan. Figure 6 illustrates a cost function for the relative accuracy of demand impacts for a residential conservation program that was estimated using existing load data and initial program design impact estimates. The expected accuracy of the estimate of coincident peak demand impact in a normal year (the X-axis in the figure) is plotted against the cost of the required evaluation. Note that obtaining higher levels of accuracy becomes increasingly expensive. This is because relative accuracy is non-linear in required sample size. At its simplest level, the method for calculating this cost function is as follows: In this case, only whole-premise load data for coincident peak demand were available. Therefore, the first step in

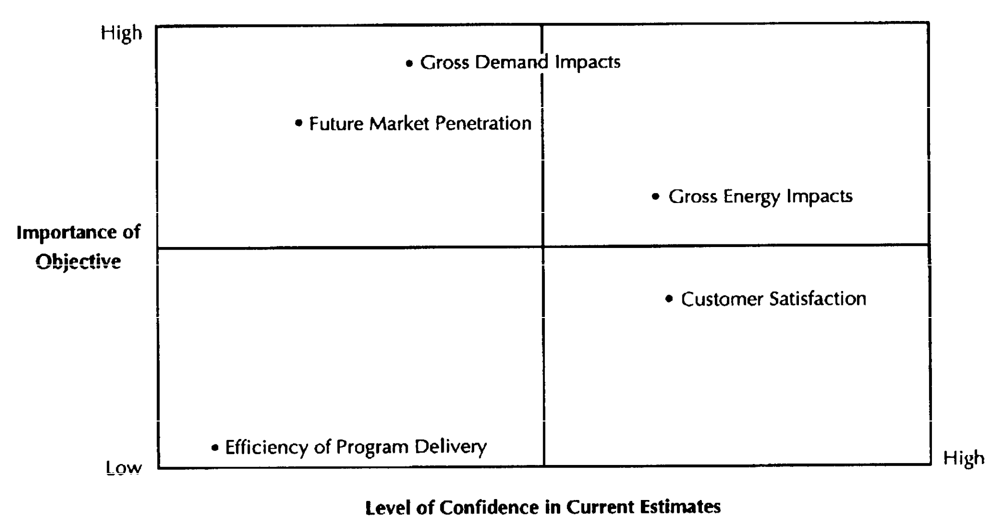


Figure 5. Importance of Evaluation Objectives vs. Level of Confidence in Current Estimates

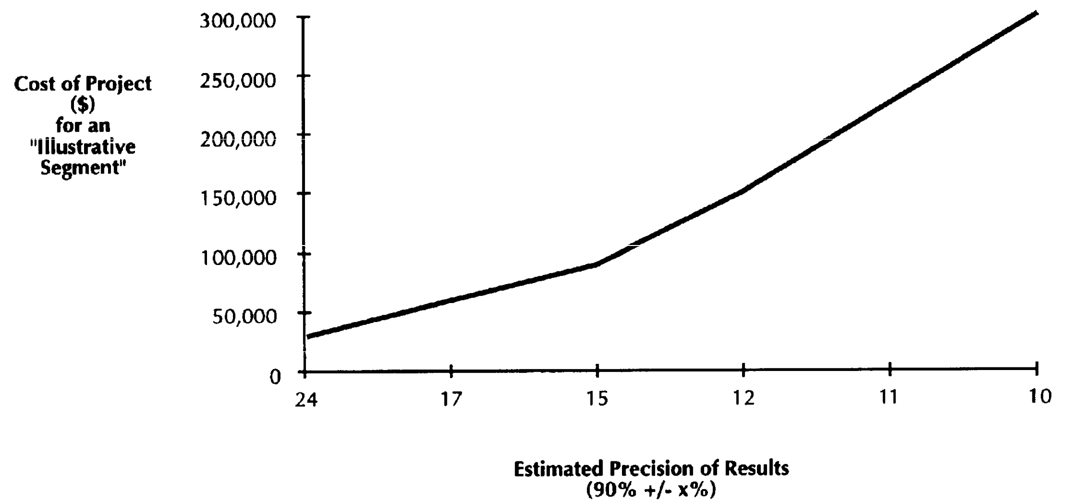


Figure 6. Project Cost by Precision of Results

the process entailed determining the sample sizes required to estimate changes in whole-premise coincident demand at the specified confidence level, from the summer before the conservation action to the summer after the conservation action. Then, the expected relative accuracy of the impact estimate is computed by dividing the accuracies achieved through the analysis of the whole-premise data (for example, 90% ± *300 watts)* obtained for various sample sizes in the previous step by the program design impact estimate. For example, if the design estimate is 900 watts, the resulting relative accuracy will be 90% ± 33%.

The appropriate calculation of relative accuracy is generally assumed to be the total variation in the demand impact estimate for a normal coincident peak hour. There are two serious problems with this assumption. (1) The utility planning objective could be inconsistent with normal coincident peak load reduction. For example, programs designed to reduce transmission and distribution (T&D) expenditures need to measure local area coincident peaks that may occur at times noncoincident with system peak. Moreover, T&D expenditures are planned based on adverse peak conditions, highest peak in 10 years, for example, not "normal peaks". And (2) as utility planners move to a level playing field, the following technical question needs to be raised: "Is the relative accuracy computed appropriate for comparing the risks of DSM accomplishments with the uncertainty of the competing supply-side resources?" Arguably, the relative accuracy computed in an evaluation may overstate the risks of DSM accomplishments relative to the competing supply-side resources.

Value of Improved Estimates for "Optimizing" Evaluation

By comparing the (marginal) cost of obtaining increased accuracy and the (marginal) value of that accuracy to the utility, decision makers can estimate the "optimal" level of evaluation; the level that maximizes the value of the usable DSM resource to the utility. In this step, the values of the different types and levels of evaluation are compared with their costs to determine how much evaluation is worth to the utility. The marginal value analysis of the type presented in Figure 7 is conducted by assessing the value to system planning of DSM resources defined at different levels of accuracy. The marginal value is obtained by running IRP models with different "types" of DSM resources, where "type" is defined as a combination of the mean value of a resource and the level of confidence in the estimates.

Example

Figure 8 illustrates how we undertook systematic evaluation planning for a U.S. utility to calculate the benefits of certain evaluation activities for their residential and non-residential programs. From the figure, the added evaluation activities centered on reducing the uncertainty surrounding kW demand impact for the Residential A/C Retrofit program is estimated to produce over $2.5 million in benefits through 2000, based on avoided costs. This benefit is estimated to cost $1.1 million over four years. Similar information for the utilities other residential and non-residential programs show that over $15 million in cumulative benefits can be obtained from enhanced evaluation that would cost an estimated $4.1 million over four

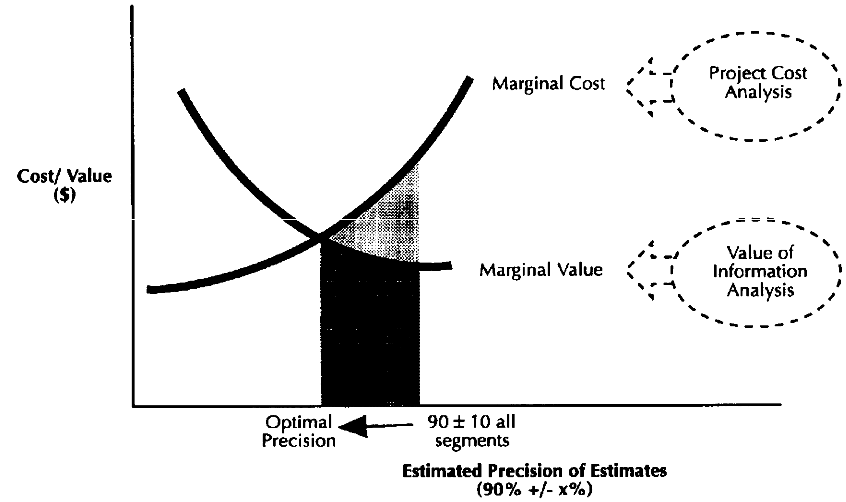


Figure 7. Analysis of Cost vs. Value of End-Use Results

PROGRAM	Participation		Current Evaluation Result or Design Estimate		Expected Evaluation Results			
	Number of Units Reported		Adjusted MW Impact	Relative**	Four Year	90% Confidence Interval		
	1992 Only	Cum to 2000	Cum 2000	Accuracy ± percent	Evaluation Budget* $(000)	Relative Accuracy ± percent	Increase in Lower Bound Cumulative 2000	
RESIDENTIAL							kW	Avoided Costs
Residential A/C Retrofit	2,340	44,534	14,653 kW	81%	1,152.5	20%	8,961 kW	$2,667,112
Attic Insulation	195	5,263	1,417 kW	81%	43.0	50%	441 kW	$35,620
Sunscreen	2,508	17,892	6,381 kW	81%	156.5	40%	2,626 kW	$683,491
A/C Load Managememt	4,811	72,948	81,702 kW	60%	583.0	20%	32,681 kW	$4,628,800
Swimming Pool Tripper	1,905	3,476	1,787 kW	80%	177.3	40%	715 kW	$303,600
Water Heater Blankets	115	7,701	308 kW	100%	14.4	80%	62 kW	$11,600
NONRESIDENTIAL								
Commercial A/C Retrofit	155	8,419	2,646 kW	80%	49.2	75%	132 kW	$30,864
Commercial Lighting	2,517	11,481	10,513 kW	80%	607.8	40%	4,205 kW	$1,774,645
Custom Rebate	2,973	22,574	22,574 kW	80%	589.1	35%	10,158 kW	$2,484,900
Commercial Audits	1,968	22,328	22,328 kW	80%	223.5	35%	10,048 kW	$2,791,800
High Efficiency Motors	2,136	120,967	2,903 kW	80%	59.3	75%	145 kW	$38,200
Total			167,211 kW		$4,159.5		70,175 kW	$15,450,632

* Total evaluation budget includes additional costs that apply to all residential and all nonresidential programs.

** Relative accuracy is statistically estimated for Residential A/C Retrofit, Attic Insulation, and Sunscreen Programs only.

Figure 8. Benefits of Evaluation Activities Reduction in Demand Impact Uncertainty

years. Figure 8 thus details the general information described above.

Conclusion

This paper suggests that a systematic economic approach to analyzing evaluation benefits and costs should be the key activity conducted in the planing of a DSM evaluation. Although application of this concept is important in being able to allocate "optimal" resources to evaluation, some caveats apply: as explained above, the evaluation accuracy estimates are always based on best available data, and often not on previous estimates of the specific objective desired; and IRP analyses should always be run under a number of scenarios; there are inaccuracies in the IRP assumptions, as well as in the DSM cost and impact estimates. Finally, the incremental costs of meeting different objectives are aggregated and presented to utility decision-makers for approval. As is the case with the demand impact estimates discussed previously, obtaining higher and higher levels of accuracy becomes increasingly more expensive for all evaluation activities. The classic economic concept of equating marginal cost and marginal value should be applied in all evaluation resource allocation activities. This ensures that the true value of evaluation is determined and that unreasonable expectations and demands are not made on the evaluators. Estimating demand impacts, for example, at 90% ± 10%, is relatively costly. Requests for such "highly accurate" evaluation results must be made with full appreciation of the associated costs. We have found when this systematic approach has been used, program evaluators, planners, and implementors have each benefited.

References

Association of Demand-Side Management Professionals. 1993. *Annual DSM Industry Report.*

Chaisson, J. M., and S. E. Coakley. 1992. "Setting Priorities for Utility Investment in Cost-Effective DSM," *ACEEE Summer Study,* pp. 8.17-22.

Kushler, M., K. Keating, J. Schlegel, and E. Vine. 1992. "The Purpose, Practice, and Profession of DSM Evaluation: Current Trends, Future Challenges," *ACEEE Summer Study,* pp. 7.1-26.

Lemaye, D., and T. Davis. 1993. "The Benefits of Demand-Side Management in a Competitive Electricity Market," *Proceedings: 6th National DSM Conference,* EPRI Report TR-102021, March 1993, pp. 319-323.

Willems, P., J. Ciraulo, and B. Smith. November 1993. *Demand-Side Management Program Evaluation and the EPA Conservation Verification Protocols,* EPRI Report TR-103103.