CALCULATING THE UNCERTAINTY IN ESTIMATES OF DSM PROGRAM COST-EFFECTIVENESS

Richard Sonnenblick, Lawrence Berkeley Laboratory, Berkeley, CA
Joseph Eto, Lawrence Berkeley Laboratory, Berkeley, CA

Introduction

Past discussions of DSM program evaluation have suggested that the appropriate level of evaluation is dependent on the cost and performance of each evaluation technique, and the value of the resulting information to the evaluator, regulator, or program planner (Refs. 1 and 2). We agree, and present this case study exploring the appropriate level of evaluation for a particular objective. Our recent LBL report has described a procedure for characterizing the uncertainties in different evaluation methods, and for relating the uncertainty of each method to the uncertainty of evaluation results, such as annual program savings, the cost of conserved energy, and program cost-effectiveness (Ref. 3). In this paper, we report on one aspect of this work, relating the precision and bias of evaluation methods to estimates of a program's cost-effectiveness.

Estimates of the cost-effectiveness of DSM are based on evaluations of program impacts. The evaluation methods used are, to an extent not well understood, subject to errors of imprecision and bias. Evaluation imprecision can reduce evaluator confidence in estimates of program cost-effectiveness, and evaluation bias can result in non-cost-effective programs being mislabeled as cost-effective. We assess the uncertainty in estimates of DSM program cost-effectiveness for evaluation methods of varying precision and accuracy. By first examining the effects of imprecision and bias, we can then assess the impact of evaluation method choice on our confidence in the cost-effectiveness of a program. The results of these calculations enable us to discuss the appropriate levels of DSM program evaluation with the objective of confidently assessing cost-effectiveness.

We begin with a discussion of key terms and a description of a framework for determining the appropriate level of evaluation. We then discuss the range of cost-effectiveness estimates observed in recent commercial lighting rebate programs. Based on this range of cost-effectiveness estimates we assess the effects of imprecision and bias on evaluator confidence in program cost-effectiveness, and discuss the implications of these findings for future evaluations.

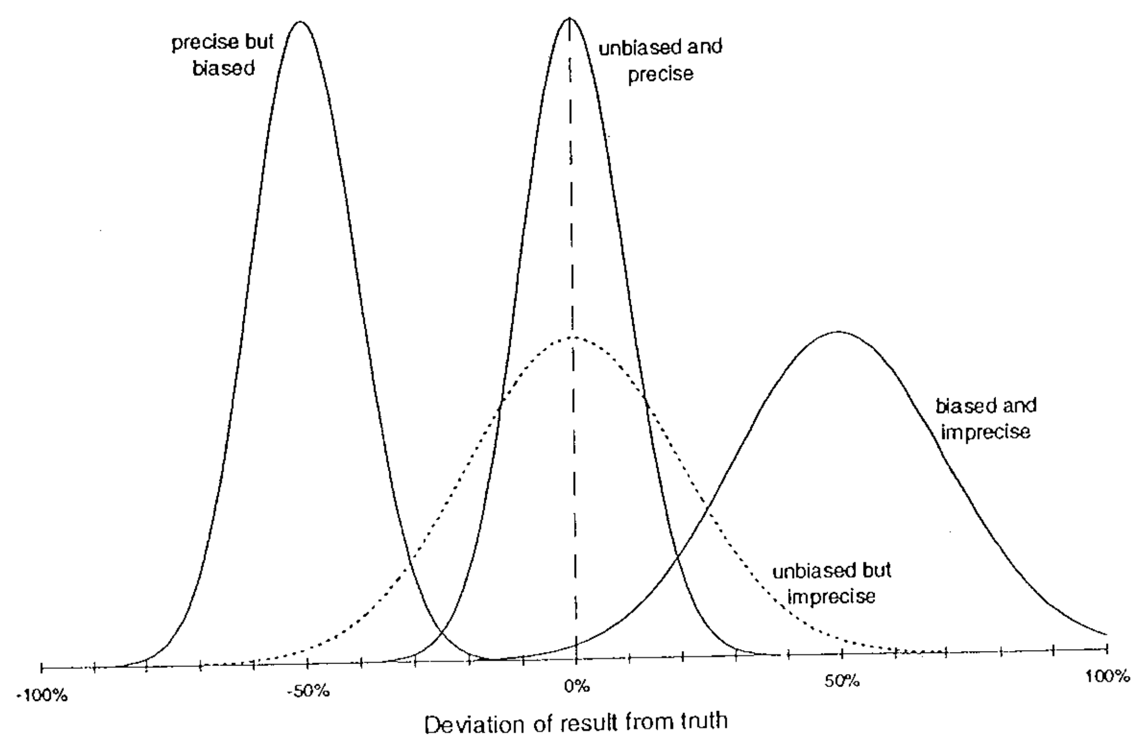

Figure 1. Bias and Precision in Savings Estimates

Uncertainty in Evaluation: Defining Precision and Bias

We use two different metrics to assess how well evaluation methods reveal a program's actual energy savings (and resulting cost of conserved energy and cost-effectiveness): bias (also known as accuracy) and precision. A biased estimate systematically deviates from the true value, under or over estimating savings. For example, if a method consistently under estimated actual savings by 20%, that method would be considered biased.

The issue of precision is more esoteric. Many program evaluations omit all discussion of estimate precision, and report savings estimates as single values. But because of the difficulties associated with calculating program savings, any estimate of program savings is subject to some uncertainty. It is this uncertainty that is measured by an expression of precision. An estimate which omits an estimate of precision is incomplete and can be misleading. For example, an estimate of annual savings of 5,000 kilowatt-hours (kWh) with a standard deviation of +/- 300 kWh is very different from an estimate of 5,000 +/- 3,000 kWh. The latter estimate is of less use as a gauge of program savings, because it suggests that the actual savings could be considerably above or below the mean estimate of 5,000 kWh, while the former estimate is more precise, satisfying what is known as a 90/10 criterion; +/-10% relative precision at a 90% confidence interval. Thus, figures reported without an estimate of this uncertainty are not as informative as those which include it.

It is important to consider the relative importance of precision and accuracy. A precise but biased estimate is worth little, unless the magnitude of the bias is known. On the other hand, an unbiased but imprecise estimate can still be useful because, on average, it provides the correct value. Figure 1 illustrates the relationship between bias and precision.

Biased, i.e., under- or over-estimates of savings, have important implications on several levels: For the utility, biased estimates of savings misinform about program cost-effectiveness. Biased over-estimates of savings may cause utilities to retain DSM programs which are not, in reality, cost-effective. At the state regulatory level, overestimates of savings will result in utility overcompensation for lost revenues (for lost revenues which, in fact, were never lost) and payment of excessive shared savings incentives. Thus, the utility is allowed to collect additional, unjustified revenue from ratepayers. At the national level, plans to reduce national dependence on fossil fuels or reduce power plant emissions using DSM activities may fall short of desired goals if plans are based on studies which exaggerate actual savings.

An *imprecise* estimate of savings has some slightly different implications: Imprecision in annual savings or measure lifetimes can affect the mean cost of conserved energy estimate, and reduce confidence that a marginally cost-effective program is really cost-effective. Most of the regulatory concern regarding precision suggests a fundamental desire for a precise estimate, but this desire is not necessarily based in the requirements of any particular use of the evaluation results. In many cases the 90/10 criteria is applied to estimates of annual savings, without a similarly rigorous criteria being required for lifetime savings or for the resulting estimates of the cost of conserved energy. Vine and Kushler, in a paper found in these proceedings, discuss the history of regulatory mandates for evaluation precision (Ref. 4). In this paper, we suggest that a precision criteria of 90/10 is usually unecessary for confidently verifying cost-effectiveness. Bias in evaluation results, depending on the evaluation methods used, appears to be a greater threat to accurate cost-effectiveness calculations.

Assessing Cost-Effectiveness

The cost-effectiveness of utility DSM programs is gauged by comparing a program's cost of conserved energy, the levelized cost of the program over the installed equipment's anticipated lifetime, to the sponsoring utilities' avoided costs.[a] A program that provides kWh savings at a levelized cost equal to or less than the levelized avoided costs is considered cost-effective, and has a total resource cost (TRC) test ratio greater than one (Ref. 5).

Even if an estimate of savings results in a TRC test ratio greater than one, the evaluator cannot rule out the possibility of the program not being cost-effective without some estimate of the savings, cost, and avoided cost estimate precision. Due to the nature of the cost of conserved energy calculation, a more imprecise savings estimate increases the probability that a program's cost of conserved energy is larger than anticipated, which can shift the mean TRC test ratio to less than one. Under certain circumstances, an imprecise estimate of savings can dramatically reduce confidence in program cost-effectiveness.

While an imprecise estimate of savings can reduce confidence in a program's cost-effectiveness, a biased estimate of savings can misrepresent a non-cost-effective program as cost-effective. Because assessment of bias requires an independent estimate of the 'true' savings for comparison, our characterization of bias is understandably less complete, but not necessarily less important, than our characterization of savings estimate imprecision.

---

a

Avoided costs are also levelized over the life of the efficiency measures using a discount rate equivalent to the utilities' cost of capital.

Table 1. Total Resource Costs and Avoided Costs from 20 Commercial Lighting Programs (Eto *et al.* 1994)

Sponsoring Utility	Annual Program Savings (GWh)	Cost of Conserved Energy (¢/kilowatt-hour)	Avoided Costs (¢/kWh)	Total Resource Cost Test Ratio
BPA	2.4	4.5¢	4.7¢	1.0
BHEC	2.1	4.7¢	5.0¢	1.1
IE	1.1	4.4¢	4.8¢	1.1
NMPC	101.4	6.0¢	9.0¢	1.5
BECo	8.3	7.2¢	11.2¢	1.6
GMP - Small C/I	3.0	7.6¢	12.1¢	1.6
PG&E	115.7	5.0¢	8.5¢	1.7
SDG&E	2.0	4.1¢	7.2¢	1.7
SMUD	43.7	6.5¢	11.2¢	1.7
CHG&E	16.1	3.7¢	6.8¢	1.9
GMP - Large C/I	16.3	6.3¢	12.1¢	1.9
SCL (Pilot)	1.1	2.5¢	4.7¢	1.9
Con Edison	91.9	6.8¢	14.0¢	2.1
NEES - Small C/I	23.5	5.2¢	10.8¢	2.1
CMP	15.7	1.8¢	4.6¢	2.5
NEES - EI	104.3	3.7¢	10.0¢	2.7
NU - ESLR	149.8	2.5¢	8.1¢	3.2
NYSEG	53.9	2.3¢	10.0¢	4.3
SCE	72.8	1.2¢	7.2¢	5.8
PEPCO	40.5	1.2¢	7.5¢	6.4

The Cost-Effectiveness of Commercial Lighting DSM

The recent DEEP Commercial Lighting Report estimated the cost of conserved energy and reported utility-estimated avoided costs for 20 commercial lighting programs (Ref. 6). Examining the ratios between the estimates of avoided costs and total resource costs for these 20 programs provides some insight regarding the distribution of typical cost-effectiveness estimates. Table 1 lists the utility estimated avoided costs, the cost of conserved energy, and the TRC test ratio for the 20 commercial-sector lighting programs examined in the DEEP report.

When point estimates of the cost of conserved energy were compared to each utilities' estimate of their avoided costs, all of the programs examined in the DEEP report were cost-effective, i.e., had TRC test ratios greater than or equal to one. A few (15%, but only 1% by energy savings) were only marginally cost-effective, with ratios less than 1.5. The majority (55%, 50% by energy savings) had cost-effectiveness ratios ranging from 1.5 to 2.1. A final group (30%, 50% by energy savings) had cost-effectiveness ratios ranging from 2.5 to 6.4. These three groups form the basis for our parameterization of cost-effectiveness estimates. We can simulate three programs with mean cost-effectiveness equal to the mean from each of the three groups.[b] We can

[b] Avoided cost calculation is a complicated matter. A complete accounting involves estimation of a utilities' fixed and variable costs per kWh and per kW supplied. These costs will vary over the life of program measures, and a thorough understanding of the utilities' resource acquisition plans is required to estimate future changes in avoided kWh and kW costs. Finally, DSM program characteristics also affect the calculation of pertinent avoided costs: A program that saves energy on-peak will have a larger avoided kW cost component than a program that only saves energy during off-peak hours.

With this in mind, it is clear that avoided cost estimation is itself subject to considerable uncertainty. While we recognize the importance of correct avoided costs calculation, an in-depth discussion of the uncertainties associated with avoided cost estimation, or of the utility and customer-borne cost elements in the cost of conserved energy, is beyond the scope of this paper.

Table 2. Parameterization of TRC Test Ratios Precision of Bottom-Up and Top-Down Evaluation Methods

Range	Mean Total Resource Cost Test Ratio	Range of TRC Test Ratios	% of the 20 DEEP Sample Programs in Range	% of Annual Savings in Range
Low	1.1	1.0 - 1.1	15%	1%
Medium	1.8	1.5 - 2.1	55%	50%
High	4.2	2.5 - 6.4	30%	49%

Table 3. Parameterizations of Annual Savings Estimate Precision

Precision	Precision from Econometric Analysis with Simulated Data	Precision from Analysis of End-Use Metering Data
Low	15%	50%
Medium	10%	25%
High	5%	10%

then estimate the effects of an imprecise estimate of savings on the cost-effectiveness estimate for each program. Table 2 summarizes our parameterization of cost-effectiveness estimates.

This section summarizes estimates of evaluation precision from our analyses of both top-down (econometric methods based on whole-premise billing data) and bottom-up (metering methods utilizing information on specific equipment installed) estimates. Myriad factors can affect the precision of both methods, and the estimates of precision given here are based on limited program data and a subset of all available evaluation methods. Thus, these estimates of precision do not universally apply to every econometric or metering study one could conduct, but rather provide a rough estimate of the range of precisions one could expect using a variety of methods.

It is also important to note that estimates of precision obtained with different evaluation methods are not strictly comparable. A value's precision is entirely dependent on the implicit assumptions that govern which aspects of a quantity are thought to be imprecise. The precision of an end-use metering-derived savings estimate is typically based on information on the sample size and sample homogeneity when compared to the participant population. The precision of an econometrically derived savings estimate is based on the capacity of the econometric model to systematically explain variability in the participant billing data. The statistical assumptions inherent in multivariate regression (e.g., normality and independence) also implicitly affect the calculation of estimate precision.

In order to create these rough estimates of the precision (and rough estimates of bias, which we discuss later in the paper) associated with different evaluation methods for commercial lighting programs, we have, in a separate report, performed a number of detailed analyses based on both actual program and simulated program data (Ref. 3). To estimate the bias and precision of end-use metering methods, we compared results from a handful of short and long-term metering studies, investigating hours of operation, sample size and selection, and interaction effects between heating cooling, and lighting equipment. To investigate the bias and precision of econometric methods we used the building energy modeling program DOE2 to simulate a set of participant and nonparticipant buildings' monthly energy consumption, and estimated econometric models using the results.

Table 3 presents the range of relative precisions (at the 90% confidence level) we obtained in the aforementioned analyses. To represent a diversity of evaluation methods, we parameterize the precision of these evaluation methods with low, medium, and high precision estimates. One should not deduce from Table 3 that econometric methods are inherently superior to metering methods. Our method of obtaining estimates of econometric precision used simulated consumption data which probably understated the variability in an actual set of monthly billing data. As mentioned earlier,

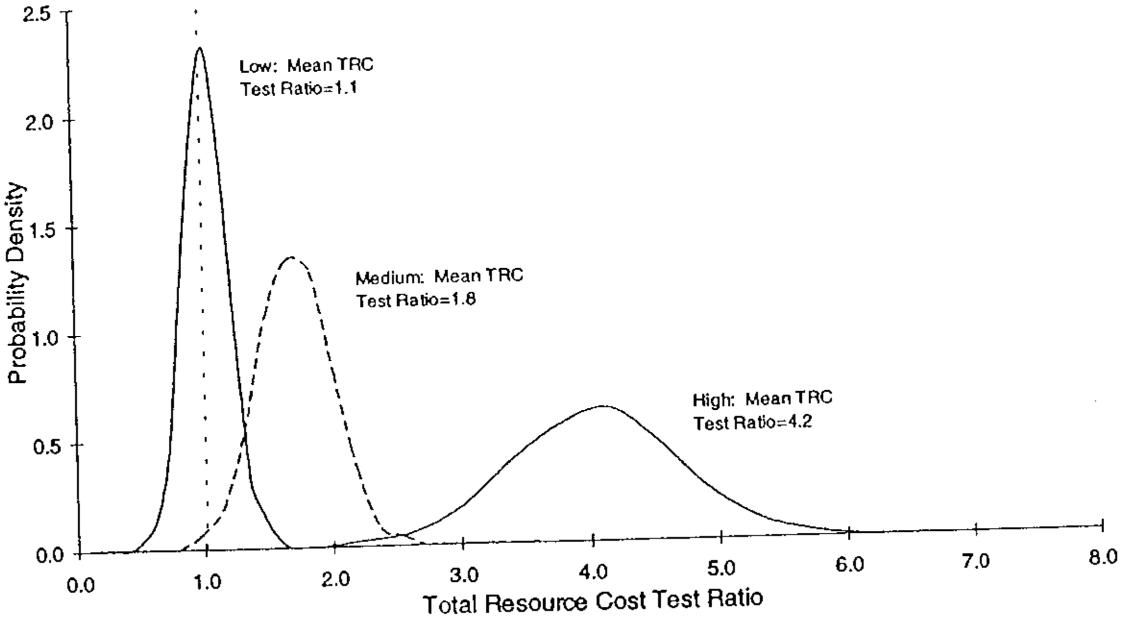

Figure 2. Distributions of the Total Resource Cost Test Ratio for Medium Precision Metering

estimates of precision from different methods are based on different statistical assumptions, and are therefore not strictly comparable. Finally, end-use metering provides a wealth of additional evaluation information above and beyond simple estimates of annual program savings.

The Effect of Imprecision on Cost-effectiveness Estimates

In this section, we use the previous sections' information on the imprecision of evaluation method results and the parameterization of cost-effectiveness to estimate the effects of imprecision on confidence in program cost-effectiveness estimates. We utilize a Monte Carlo model to propagate uncertainties because the method and results are easily grasped without a detailed understanding of calculus or other analytic propagation of error techniques, and because Monte Carlo techniques allow more freedom in specification of uncertain quantities and functional relationships.

Additional uncertainty is incorporated into the cost-effectiveness calculation with the incorporation of an uncertain measure lifetime estimate, based on inventories of efficient equipment installed in lighting programs in the Pacific Northwest (Ref. 7). Most regulators focus on the precision of annual savings. By incorporating an uncertain estimate of measure lifetime, we can estimate the precision of lifetime savings and program cost-effectiveness.

Monte Carlo Model Results

Three examples of the resulting distributions of the TRC test ratio from the Monte Carlo model are given in Figure 2.[c] The distributions displayed reflect annual savings estimates of average precision obtained through end-use metering. Each of the three distributions represents a different mean estimate of the TRC test ratio, representing the three parameterizations described in Table 2.

The distributions with some portion of their area to the left of 1.0 represent programs which, given the precision of the evaluation methods used, could be non-cost-effective even though the mean estimate, which might be submitted alone as an estimate of cost-effectiveness in a regulatory hearing, is greater than 1.0.

Table 4 lists the fraction of each distribution that lies below 1.0, indicating the likelihood of non-cost-effectiveness. Only the distributions for programs with low mean total resource costs have a significant portion of their area below 1.0. Thus, the risk of mistakenly labeling a program cost-effective when it actually is not is highest for programs whose mean estimates of the TRC test ratio are close to 1.0. This

[c] The Monte Carlo model sampled 1000 points from each distribution, obtained using median hypercube sampling.

Table 4. Fraction of Distributions Representing Non-Cost-Effective Programs

Mean TRC Test Ratio	Savings Estimation Method	Precision	Percent of Distribution Less Than 1.0
Low (1.1)	End-Use Metering	Low (±50%)	40%
		Medium (±25%)	29%
		High (±10%)	11%
	Econometric	Low (±15%)	19%
		Medium (±10%)	11%
		High (±5%)	3%
Medium (1.8)	End-Use Metering	Low	7%
		Medium	—
		High	—
	Econometric	Low	—
		Medium	—
		High	—
High (4.2)	End-Use Metering	Low	1%
		Medium	—
		High	—
	Econometric	Low	—
		Medium	—
		High	—

result is intuitive: imprecise measurement which results in a ratio close to one has a greater chance of actually being below one than a similarly imprecise measurement which results in a ratio much larger than one.

Implications of Estimate Imprecision

The Monte Carlo results summarized in Table 4 have important implications for the level of precision required to confidently assess DSM program cost-effectiveness. The answer to the question, "Is a 90/10 criterion necessary to confidently assess the cost-effectiveness of a DSM program?" is "No". Only for programs with mean TRC test ratios near 1.0 is a level of precision approaching 90/10 necessary to confidently determine whether the program is truly cost-effective. Even for the lowest precision evaluation, a program with a 'medium' mean TRC test ratio is cost-effective at the 90% confidence level.

Should these results change the way in which evaluations are conducted? We see two ways to proceed from this analysis: In the distribution of TRC test ratios from the DEEP sample of 20 commercial lighting programs, we observe that the majority of programs fall into the 'medium' category. Thus, in the majority of cases, a 90/10 criterion would be excessive for the determination of cost-effectiveness. It follows that a less stringent precision requirement should be adopted.

Alternatively, program planners and evaluators may have some previous estimate of the mean TRC test ratio associated with the program, perhaps based on a previous year's evaluation, or on program planning estimates. A determination of evaluation requirements could be made based on this estimate of cost-effectiveness: programs with preliminary cost-effectiveness ratios near 1.0 would be allocated additional evaluation resources to ensure a confident assessment of ex ante cost-effectiveness.

If cost-effectiveness verification were the primary goal of an evaluation, we would advocate a combined approach, whereby programs without preliminary or planning estimates of cost-effectiveness are allocated enough evaluation resources to assess cost-effectiveness for a program with a TRC test ratio in the 'medium' range, while programs with some cost-effectiveness information would be evaluated as dictated by these ratios.

A fundamental hurdle in this type of evaluation planning is our inability to estimate the cost of attaining a given level of evaluation precision. The programs for which we have been able to collect detailed evaluation data in our research represents too limited a sample to conclusively characterize the program attributes, participant characteristics, and evaluation method uncertainties required to understand the precision-evaluation tradeoff. Thus, the most practical and immediately applicable result of our analysis here is that a 90/10 criterion for relative precision of

Table 5. Sources of Bias in Cost of Conserved Energy Estimates

Parameter	Source of Bias	Magnitude of Bias in the Cost of Conserved Energy
Bottom-Up Savings Estimates	Seasonality of Hours of Operation	-5% to +5%
	HVAC/Lighting Interactions	+5% to +15%
	Nonrepresentative Metered Sample	unknown
Top-Down Savings Estimates	Engineering Estimate Uncertainty in SAE Models	+5% to +50%
Measure Lifetime	Use of Mfr. Estimates	-40% to -5%
Free Riders*	Free Riders Over Time	at least -11%
Free Drivers	Omission of Free Driver Savings	positive, but unknown

*Free riders are relevant only for utility cost test ratios, not TRC test ratios.

annual savings estimates is almost always excessive for determining cost-effectiveness.

The Effect of Bias on Cost-Effectiveness Estimates

Thus far, we have focused on the importance of precision in assessing cost-effectiveness. However, it is crucial, and potentially more important, to consider the role of bias as well. Despite the importance of estimate accuracy, our understanding of evaluation bias is less developed than our characterization of precision due to the difficulty of characterizing bias, which requires an independently estimated, unbiased estimate for comparison (i.e., the true, actual program savings). Our limited sample of program evaluations also hindered a more thorough characterization of evaluation method bias.

Just as imprecise savings estimates pose the greatest threat to programs with mean cost-effectiveness near one, those same programs may actually be non-cost-effective programs with biased estimates of savings. Table 5 reviews the biases identified in our research (Ref. 2). The biases in Table 5 are given as percentage deviations from the unbiased value. A negative bias means that the cost of conserved energy is underestimated and a positive bias means that the cost of conserved energy is overestimated.

The effect of these biases is multiplicative; a cost of conserved energy estimate based on limited duration metering and manufacturer estimates of measure lifetimes would be subject to cumulative biases which could double or halve the cost of conserved energy. Some utilities implicitly acknowledge the bias inherent in their cost-effectiveness estimates by only implementing and continuing programs with a TRC test ratio significantly above 1.0, using, for example, a threshold of 2.0 or higher to screen programs.

For TRC test ratios close to one, even a bias of a few percent could result in a non-cost-effective program being erroneously labeled cost-effective (or vice-versa, labeling a cost-effective program as non-cost-effective). However, when considering bias and precision together, the effect of bias is even more pervasive. A negative bias in the cost of conserved energy means that the distribution of the true TRC test ratio in Figure 2 is further to the left and closer to 1.0 than the supposed distribution. For programs with mean TRC test ratios close to one, a larger fraction of the distribution would move below the cost-effectiveness threshold of 1.0, revealing an increased probability that the program is not cost-effective. If these biases are large enough or several negative biases are applicable, even a program with a mean TRC test ratio in the 'medium' range could, in actuality, be non-cost-effective. For example, a bottom-up metering study could meter equipment in the winter, overestimating annual hours of operation by approximately 5%, and savings could be coupled with biased manufacturer estimates of equipment lifetimes, overestimating lifetimes by as much as 40%. In the worst case, the combined bias could overestimate lifetime program savings by 45%, which would cause a marginally non-cost-effective program to appear to have a (biased) TRC test ratio approximately equal to 1.6.

Most of the evaluations for the 20 programs reviewed in Table 1 are subject to at least one of the biases listed in Table 5: (1) Metering studies that did not adjust for seasonality or interaction effects; (2) SAE models that used imprecise tracking database estimates of savings; (3) Measure lifetime estimates based on manufacturers' estimates of equipment operation, and; (4) relevant only for utility cost test ratio estimation. Free ridership estimates which only discuss free riders in the first program year.

Given the potential importance and pervasiveness of these biases in the current practice of evaluation, it seems prudent that some evaluation resources should be allocated to reduce bias, and not only imprecision, in the cost of conserved energy. In the next section, we discuss the potential costs of reducing these biases.

Table 6. Estimates of the Cost of Addressing Biases in Commercial Lighting Evaluation

Source of Bias	Method Used to Reduce Bias	Approximate Marginal Cost
Seasonality of Hours of Operation	Seasonality Adjustment Longer Term Metering	Low Med/High*
HVAC/Lighting Interactions	Metering of HVAC Equipment Modeling of HVAC/Lighting in Prototypical Buildings	High Low/Med
Nonrepresentative Metered Sample	Proper Participant Stratification and Selection of Equipment to Meter	Med Low
Eng.Est. Uncertainty in SAE Models	Switch to non-SAE model	Low
Use of Mfr. Estimates of Lifetimes	Verify Continued Operation with Site Surveys	Med*
Free Riders Over Time	Analyze Equipment Sales to Nonparticipants During Life of Program Equipment	Med*
Omission of Free Driver Savings	Customer and Vendor Surveys Analyze Equipment Sales in Diffusion Framework	Low Med/High*

* These methods require considerable additional time for the compilation of sufficient data.

Implications of Estimate Bias

The preceding discussion demonstrates the significance of evaluation bias when assessing program cost-effectiveness. How can these biases be handled in the program evaluation? Ideally, the costs and potential impacts of each bias would be compared, and resources would be spent to identify and reduce the largest biases at the least cost. Because of the variability associated with the impacts of the biases, it is difficult to definitively prioritize the biases in order of their importance so that they can be addressed effectively given available resources. A larger sample of program evaluation data than is presently available is required to better characterize each evaluation method's biases. To begin to prioritize the treatment of the biases in evaluation, we present some qualitative estimates of the evaluation costs associated with reducing the biases.

Even with only rough guidelines regarding evaluation costs, we can draw some conclusions. Many of these biases can be at least partially addressed with minimal additional evaluation resources: Metered samples can be adjusted to control for seasonal effects and carefully stratified based on equipment, facility, and building zone characteristics; SAE models can be used only when tracking database estimates are of sufficient precision; and customer and vendor surveys can be used to obtain first-order estimates of free driver and spillover effects. Incorporating these changes into evaluation practice would improve the accuracy of estimates of cost-effectiveness of lighting programs at minimal additional cost. For those evaluation improvements which require substantial commitments of time and money, a decision analytic framework, delineated in the next section, could be used to determine if the preliminary estimate of the TRC test ratio warranted additional efforts to reduce estimate bias. The generalizability of information regarding these biases may also represent a justification for additional evaluation: If information regarding a bias from a particular evaluation can be used to estimate the magnitude of the same bias for other programs and evaluations, the cost of the additional evaluation is effectively spread among multiple programs.

As with estimate imprecision, we find that evaluation biases threaten the cost-effectiveness of programs with TRC test ratios closer to 1.0. Unlike imprecision, however, biases could threaten claims of cost-effectiveness for programs with TRC test ratios in the medium range (~1.8) as well. When imprecision is considered in addition to bias, reduced statistical confidence in even higher TRC test ratios may result.

The Value of Correctly Assessing Cost-Effectiveness

In this paper, we've discussed the role of precision and bias in assessing cost-effectiveness. The next logical step is to devise a method for the optimal allocation of evaluation resources. Ultimately, the cost of improving the precision and accuracy of evaluation method results should be traded-off against the value of obtaining increasingly accurate and precise estimates of program cost-effectiveness. A common use of cost-effectiveness information is program screening: ongoing programs are screened to determine if they should be funded for the next program year. In the following paragraphs, we briefly outline a procedure for trading off evaluation cost and evaluation value. We present this decision analytic approach as an intriguing topic for future research.

A decision analytic approach to determining the appropriate level of additional evaluation to reduce imprecision and bias requires: (1) a subjective estimate of the chances that the program is actually non-cost-effective given any initial evaluation results, and (2) an estimate of the resources that would be (potentially) misallocated to the program in the following year (i.e., next year's program budget).[d] The product of these two values is the expected value of future misallocated resources, and would represent the maximum marginal evaluation expenditure justified if the resulting evaluation provided an estimate of cost-effectiveness with 100% certainty (known as the expected value of perfect information). An additional estimate of the results of future evaluations (i.e., the certainty of the subsequent estimate of cost-effectiveness) could be used to refine the justifiable marginal evaluation expenditure. In a utility portfolio of programs, application of this technique for each program would result in evaluation expenditures which, on average, would minimize the sum of evaluation expenditures and future misallocated program resources.

Using cost-effectiveness estimates for program screening and budgeting is just one example of how savings estimates are used and evaluation resources should be apportioned. Additional applications of evaluation information such as shared savings calculation, lost revenue recovery hearings, and load forecasting may justify more accurate and precise evaluation results, and may require a different selection framework. For example, when considering shared savings incentives earned by the utility, evaluation expenditures may be justifiably apportioned to programs with high TRC test ratios, because these programs can potentially provide the utility with the largest monetary rewards[e], as opposed to cost-effectiveness screening, where programs with lower TRC test ratios would justify increased evaluation resources. The appropriate level of evaluation expenditures could be set and justified by considering the value of evaluation using one, several, or all, of these applications of evaluation results.

---

[d] A similar method is used to investigate utility planning uncertainties by Hobbs and Maheshwari (Ref. 8).

[e] High TRC test ratios result in larger incentive awards with all other things (e.g., program size) being equal. A seconary effect, where larger programs usually have higher TRC test ratios and therefore larger shared savings incentives, also exists.

Conclusions

In this paper we describe and implement a framework to assess the effects of bias and imprecision on estimates of program cost-effectiveness. The framework allows program evaluators and program planners to explicitly handle the uncertainties inherent in the complex evaluation of a DSM program. By estimating the effects of these uncertainties on estimates of program cost-effectiveness, program planners can ascribe confidence to their results and adopt levels of evaluation expenditures which are justified by the uses of the evaluation results.

Our implementation of this framework suggests that imprecision in the cost of conserved energy is significant for programs with mean TRC test ratios close to one, while higher ratios guarantee cost-effectiveness even with considerable estimate imprecision. A 90/10 criteria for precision seems excessive for most programs when screening for cost-effectiveness, in light of these findings.

However, bias in savings estimates can threaten the confidence of cost-effectiveness estimates for programs with ratios approaching 2.0, especially when estimate imprecision is also considered. Much of the contemporary concern with precision should be redirected to examine bias in evaluation estimates, given the results we present here.

Savings estimate biases and imprecision stem from a multiplicity of factors, some of which require expensive additions to evaluation procedures, and some of which require only slight changes in evaluation methods. While we recommend that all evaluations should include the least-cost methods to reduce estimate bias, additional expenditures should be traded off against the value of accurately assessing cost-effectiveness. The value of other evaluation information applications, such as demand forecasting and program improvement, require additional, explicit tradeoffs between information value and evaluation costs.

Reference

1. Hummel, Phillip E. (1993), "Resource Allocation and DSM Program Evaluation Planning", *Proceedings of the 1993 Energy Program Evaluation Conference*, Chicago, IL, pp.637-642, August.
2. Wirtshafter, Robert, Les Baxter (1991), "Establishing Priorities for Future Evaluation Efforts", *Proceedings of the 1991 Energy Program Evaluation Conference*, Chicago, IL, pp.137-142. August.
3. Sonnenblick, Richard, Joseph Eto (1995), "A Framework for Improving the Cost-effectiveness of DSM Program Evaluation", Lawrence Berkeley Laboratory, Berkeley, CA, LBL-37158, May.
4. Vine, Edward L., Martin Kushler (1995), "The Reliability of DSM Impact Estimates", *Proceedings of the 1995 Energy Program Evaluation Conference*, Chicago, IL.

5. California Public Utilities Commission and California Energy Commission (1987), "Economic Analysis of Demand-Side Managment Programs." *Standard Practice Manual*, P400-87-006, December.
6. Eto, Joseph, Ed Vine, Leslie Shown, Richard Sonnenblick, Chris Payne (1994), "The Cost and Performance of Utility Commercial Lighting Programs", Lawrence Berkeley Laboratory, LBL-34697, May.
7. Skumatz, Lisa S., Curtis Hickman (1994), "Effect of Energy Conservation Measures and Equipment Lifetimes in Commercial Buildings: Calculation and Analysis", Proceedings from the 1994 ACEEE Summer Study, Asilomar, CA, v.8, pp.193-204.
8. Hobbs, Benjamin, Prakesh Maheshwari (1990), "A Decision Analysis of the Effect of Uncertainty upon Electric Utility Planning", *Energy*, 15:9, pp.785-801.