Uncertainty in Site Inspection and Tracking Database Estimates of Savings

Richard Sonnenblick and Joseph Eto, Lawrence Berkeley Laboratory

We systematically analyze impact evaluation results of three commercial lighting rebate DSM programs. Our research includes (1) analysis of *ex ante* and *ex post* estimates of program performance, broken down into critical program parameters: hours of operation, watts saved per measure, and measures installed per site; (2) construction of probability distributions of program performance, both in the aggregate and for these critical program parameters; and (3) use of these analyses and distributions to draw conclusions about the accuracy of savings estimates from a variety of evaluation methods.

Our analysis suggests that realization rates (a ratio of metered savings estimates to tracking database savings estimates) for the sample of participants we examine are subject to tremendous variability, calling into question the usefulness of a point estimate of the realization rate. Discrepancies in estimates of hours of operation are responsible for most of the uncertainty in the realization rate. Finally, the impact of shorter measure lifetimes on savings estimates suggest that persistence studies should be an evaluation priority.

Introduction

The historical record shows wide variation in the performance of conservation programs, in part because the evaluation of these programs is both difficult and expensive. Despite the thousands of programs implemented in the last 15 years, no clear methodology exists for benefiting from the lessons learned in previous programs. Nadel and Keating's paper on post-program savings estimates sparked a debate on the need for evaluation (Nadel and Keating 1991). Our research expands on this debate by investigating the systematic biases and uncertainty inherent in tracking database savings estimates (usually known as engineering estimates), and measured consumption savings estimates (usually known as post-program measured estimates).

Past analyses of tracking database and measured consumption evaluation methods have focused only on the savings estimates resulting from each method, and then only on point estimates of these savings estimates. In this study, we examine the parameters comprising each savings estimate, the underlying variability associated with these parameters and the precision of the parameters when different evaluation methods are used to estimate them.

We restrict our analysis in this paper to three evaluation methods: engineering algorithms used in program tracking databases, engineering algorithms augmented with on-site inspection data, and end-use metering. All three methods provide program parameter estimates which can be used to calculate energy savings. Tracking database estimates of savings are the simplest, while on-site estimates and end-use metering estimates successively improve and build upon the information in the tracking database. Thus, we are able to compare directly the three methods of estimating energy savings.[1]

Numerical techniques allow us to examine the results of evaluation studies in several interesting ways (Morgan and Henrion 1990):

1. We can compute the effect of changes in parameters on estimates of savings by conducting a *sensitivity analysis.* Because our model of savings is a simple multiplicative model, model sensitivities are straightforward.

2. We can estimate the uncertainty of savings estimates induced by the uncertainty of the parameters used by the different evaluation methods, i.e., *uncertainty propagation.* We begin by examining the parameters individually, and then investigate their cumulative effects on estimates of savings.

3. We can compare the importance of the parameter uncertainties in terms of their relative contributions to uncertainty in the savings estimate, i.e., *uncertainty analysis.* This type of analysis reveals which parameters' values must be made more accurate in order to improve the precision of the savings estimate. We also describe the effects of different measure lifetime assumptions on savings estimates.

It is useful to distinguish between uncertainty associated with evaluation performance and uncertainty associated with program performance:

- *Evaluation Performance.* Shortcomings or errors in the methods used to calculate parameter values are systematic errors which represent deviations from "true" parameter values or energy savings. Systematic bias can be reduced by utilizing more rigorous evaluation methods. In this paper, we examine the systematic bias in tracking database and site inspection evaluation methods. The variation of this systematic bias across different customer sites is also important: A systematic bias with a small standard deviation can be controlled for by adjusting the parameter values used to calculate savings. A bias with a large standard deviation is less systematic, and more difficult to control for by simply adjusting parameter values.

- *Program Performance.* The "true" parameter values and resulting energy savings are inherently variable due to differences in customers, buildings, measures, program characteristics, and weather. Differences in savings achieved by different utilities implementing similar programs are due to this type of variability. Variability of energy savings estimates can be reduced by grouping savings estimates by building type, customer size, measure, etc. The data set we use in this paper does not provide measure-level data or results by building type; thus, we are unable to investigate the variability of program performance.

An Energy Savings Model for Evaluation Method Comparison

In order to analyze the parameters used to generate energy savings estimates, we need to specify an equation for program savings based on the parameters collected during evaluation. The equation we use is given below.

$$EnergySavings = \frac{measures}{site} \times \left(\frac{Watts}{measure_{old}} - \frac{Watts}{measure_{new}} \right)$$

$$\times\ Annual\ Hours \times Economic\ Measure\ Life$$

Lifetime energy savings for a measure, or set of measures, is the product of the change in wattage between the old, inefficient and new, efficient measures, the annual hours of operation, and the measure life (weighted by the change in watts for each measure if evaluating a set of measures). Our analysis is restricted to the site-level; given the data available, this is the most disaggregate level at which distinctions between evaluation methods' estimates of savings can be examined.

A peak load savings equation would require the addition of a diversity factor in order to estimate demand savings as well. While a probabilistic analysis of demand savings estimates is crucial for calculating the value of evaluation results in DSM and capacity planning exercises, this paper focuses on energy savings estimates.

Evaluation Methods Examined

The following paragraphs describe the evaluation methods used to collect program information used in our analysis.

Tracking Database Estimates. Tracking database estimates of savings use estimates of hours of operation and watts per measure based on building and measure types. The tracking database provides census information on the number of program measures installed.

Simple tracking estimates can be augmented with information from customers on estimated hours of operation for their site and the type and efficiency of measures being replaced with efficient equipment through the program. Adding this information tailors the savings estimates to each site, but since the information is compiled through phone or mail-in surveys with participants, and are not verified by trained utility personnel, the resulting parameter estimates are still subject to errors.

On-site Inspection Estimates. On-site inspections involve visits to customer facilities by trained utility inspectors who gather information on efficiencies of equipment to be replaced, hours of equipment operation, the actual number of measures installed, and the present condition of program-installed equipment. However, on-site inspections do not verify the actual rate of equipment's electricity consumption, relying instead on engineering tables of equipment consumption. In addition, hours of operation are deduced from facility hours, from short term observation of equipment, or from discussions with building occupants.

End-use Metering Estimates. End-use metering is usually performed for at least a month prior to and a month after installation of efficient equipment. Meters connected directly to equipment installed by the program measure actual hours of operation and kW consumption over time. Data loggers are not 100% accurate, and metered results are subject to error. The expense of end-use metering usually requires that a subset of all measures installed be meters, which can lead to sampling errors if the measures metered are not representative of the general population.

Program Data

While the quality of impact evaluations being performed is constantly improving, only a small handful of evaluation reports present enough data to examine changes in program parameters across evaluation methods. We found only three programs with the required data in the area of commercial lighting DSM. All three programs included similar measures. Evaluation methods used in all three programs included site inspections and surveys, short duration metering, and spot watt metering. The three programs are New England Electric System's 1991 Energy Initiative and Small C&I programs, and Northeast Utilities' 1991 Energy Saver Lighting Rebate program (RLW Analytics 1992, 1992a, 1992b).

Since this analysis involves an aggregation of information from several programs, the results are not representative of these three programs individually, and should not be generalized to all commercial lighting programs. Our analysis of these ~80 customers in three programs should be used as a case study which demonstrates how to use mathematical models to investigate uncertainty.

Realization Rates

Recently, some evaluation analysts have begun to use the term "realization rate" to explain differences between tracking estimates of savings and their final savings estimate based on extensive *ex post* evaluation. We use this term here with several qualifications:

No single parameter estimation method can provide the absolute truth in terms of a program's kWh savings. While tracking databases are most susceptible to misestimation of savings, even end-use metering can misestimate savings due to sample size or sample representativeness limitations, and meter mechanical/electronic difficulties. Thus, no evaluation technique is capable of providing unqualified estimates of the savings "realized" by a program.

Similarly, every realization rate is unique because every evaluation is unique. If tracking database estimates of savings (the denominator of the realization rate) include adjustments for free riders, hours of operation surveys, etc., then one should expect concomitant changes in the realization rate. Differences in post-program evaluation (the numerator of the realization rate) can also alter the realization rate. Comparing realization rates from different programs, even if one knows the realization rates with extraordinary precision, may be like comparing apples and oranges if differing evaluation methods are used. See Eto et al. (1994) for some examples of how differences in evaluation methods affect the resulting realization rates.

The realization rates, based on the ratio of the point estimate of end-use metering savings to the point estimate of tracking savings, are given in Table 1. Note that these realization rates are obtained using tracking database savings estimates in tandem with end-use metering, and are different than those obtained using regression models, such as an SAE analysis. They are also different than the realization rates computed by Nadel and Keating, which compared post-program savings estimates with pre-program savings estimates.

Table 1. Original Realization Rates

Program	Realization Rate
NU ESLR	.87
NEES EI	.70
NEES Small C&I	.88

While previous studies consider realization rates as an end result, these realization rates are the starting point for our more detailed investigation and should not be considered separate from the paper's analysis. Forthcoming sections examine the savings equation parameter values' uncertainty in order to understand what the realization rates represent.

Systematic Errors in Savings Parameters

End-use metering and site inspections provide site and measure-level data. These data allows us to compare the values of parameters used in engineering algorithms with the parameter values obtained in the field. In this section, we examine the mean and standard deviations for the parameters used as inputs to the savings model.

Rather than examine the parameter values themselves, we chose to examine ratios of the parameter values obtained using different methods. The ratios describe the extent to which parameter values differ when different evaluation methods are used. Table 2 provides estimates of measures installed per site, hours of operation, and watts per measure, obtained using end-use metering and site inspections, for several programs. The numbers presented in Table 2 are expressed as a ratio of the parameter value obtained using metering to the value in the program's tracking database. For example, the tracking database underestimated the number of measures per site, on average, by 3% for NEES' Small C&I program.

While the number of measures installed per site are underestimated slightly in the tracking database, the tracking database overestimates every other parameter. All but one parameter (hours of operation in the Energy Saver Lighting Rebate program) is overestimated through site inspections. The tracking database overestimates the actual savings per site by overestimating the individual parameter values used in the equation to calculate savings. Multiplying the parameter level realization rates yields the basic realization rates presented in Table 1.[2]

Even though the parameter values in Table 2 suggest the existence of a systematic bias in the tracking estimates, it is equally important to examine the variability of this bias. This is different than simply examining the variability in a single parameter, such as hours of operation, across sites. Here we are interested in the variability of the *ratio of* tracking database estimates and metered estimates (or site-inspection estimates) for a parameter. A small variability would indicate that adjusting parameters in the tracking database could dramatically improve tracking database accuracy and subsequent estimates of savings. But a large variability would suggest that important, extraneous factors could be missing from the parameter values used in the tracking database, and deciding on an accurate adjustment factor for tracking database estimates would be difficult.

Examining the ratio of savings estimates for the NEES and NU programs in our sample reveals a significant variability. We illustrate this variability in Figure 1 by plotting the ratio of metered parameter values to tracking parameter values, and of metered parameter values to site inspection parameter values, along with each ratio's standard deviation.

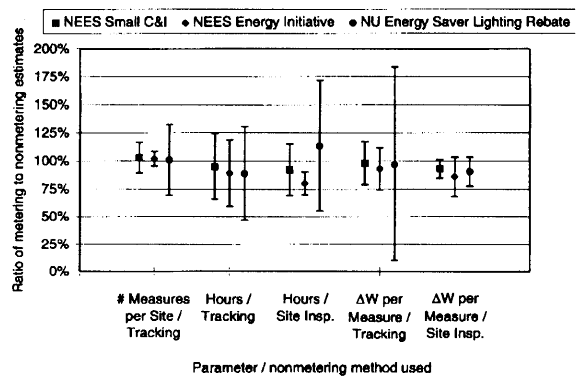

Figure 1. Differences Between Parameter Values

The systematic bias described in Table 2 is framed by a much larger stochastic component, as illustrated in Figure 1. The large variability in parameter values obtained with different methods may mean that the parameter values used in the tracking database, while fairly accurate on average, are inaccurate for a large number of individual sites and/or measures. The greater the stochastic component, the more difficult it is to generalize from the metered sites to a larger sample of participants, and the more. difficult it becomes to systematically correct for error by adjusting tracking database estimates.

Of particular interest are the wide variations in NU's Energy Saver Lighting Rebate program between site inspection and metered estimates of hours of operation, and between tracking and metered estimates of the change in watts per measure. Evaluators found inaccuracies in the tracking database algorithms used to calculate the change in watts for optical reflector retrofits and metal halide retrofits. These errors in the tracking database calculations can explain the large standard deviation for the change in watts parameter: savings from metal halide retrofits were systematically underestimated and savings from optical reflector retrofits were systematically overestimated. However, no reason was given in the evaluation reports for the discrepancy between hours of operation estimates based on site inspections, and those based on data loggers.

Table 2. Comparison of Parameter Values from Different Evaluation Methods

Ratio of Metered Estimate to:	# Sites Metered	Measures per Site	Hours of Operation		Watts Saved per Measure	
		Tracking	*Tracking*	*Site Insp*	*Tracking*	*Site Insp*
NEES Small C&I	21	103%	92%	96%	93%	96%
NEES Energy Initiative	23	102%	80%	89%	86%	93%
NU Energy Saver Lighting Rebate	30	101%	89%	113%	97%	91%

In the next section, we describe the distributions we use in place of each program parameter and perform uncertainty and sensitivity analyses on the energy savings equation.

A Probabilistic Model of Energy Savings

Recalling our basic energy savings equation for lighting measures:

$$EnergySavings = \frac{measures}{site} \times \left(\frac{Watts}{measure_{old}} - \frac{Watts}{measure_{new}} \right)$$

$$\times\ Annual\ Hours \times Economic\ Measure\ Life$$

we can use probabilistic values based on information collected in site inspections and end-use metering in place of the point estimates usually used for A *Watts*, and *Annual Hours.* We can analytically specify distributions for measure lifetimes based on the limited persistence information available. Then, using Monte Carlo methods to estimate the equation, we can obtain a probabilistic estimate of energy savings from the model. The distribution of the resulting estimate represents the underlying uncertainty associated with energy savings for the program (or programs) from which the probabilistic estimates of each parameter were calculated.[3]

For this part of the analysis, we use the site inspection and metering data from NEES Small C&I and Energy Initiative programs to construct probability distributions for the number of measures per site, hours of operation, and watts per measure for the model. We construct three different sets of input distributions:

The first set of input distributions will be based on the differences in parameter values obtained using *end-use metering* and those in the *tracking database.* The resulting outcome distribution will express the extent to which savings estimates obtained using end-use metering differ from estimates in the tracking database. If end-use metering results are assumed to represent actual savings, then the outcome distribution generated here represents the degree to which tracking database savings estimates deviate from this reality.

The second set of input distributions will be based on the differences in parameter values obtained using *site inspections* and those in the *tracking database.* The outcome distribution estimated using these parameters will express the difference between site inspection estimates of savings and tracking database estimates.

The final set of input distributions will be based on the differences in parameter values obtained using *end-use metering* and those obtained with *site inspections.* The outcome distribution estimated using these parameters will describe the variation of site inspection estimates of savings from end-use metering estimates, and can be interpreted to be the degree to which site inspection estimates of savings deviate from this reality.

As a first approximation, parameters in our sample can be approximated with a normal distribution.[4] For example, a histogram of the difference between tracking database estimates of hours of operation and metered estimates of hours of operation from the NEES programs are plotted in Figure 2.

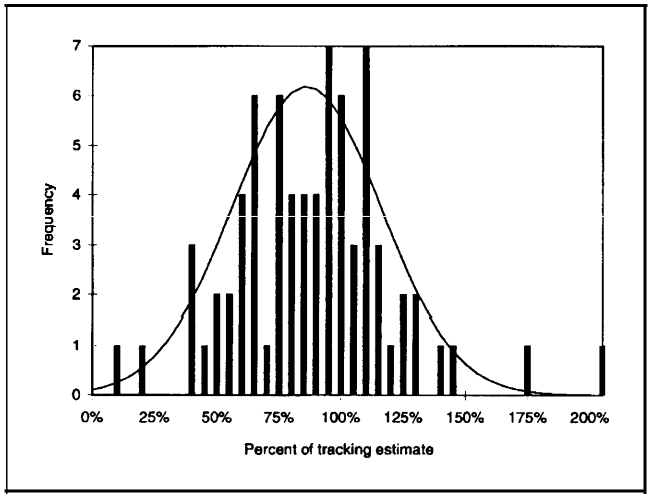

Figure 2. Distribution of Hours of Operation Realization Rates

Rank transformations were performed on each set of parameter data to verify the fit of a normal distribution. Latin Hypercube sampling was used to obtain outcome distributions, computing 1000 sample points per set of input distributions.

Uncertainty Propagation

The average and standard deviations of the outcome distributions from all three sets of input distributions are given in Table 3.

If end-use metering most closely approximates the actual energy savings for the sample, then tracking estimates of savings overestimate energy savings, on average, by approximately 22% and savings estimates based on site inspection data overestimate energy savings by approximately 12%.

While one may be tempted to conclude that the 78% figure in Table 3 is a transferable ‘realization rate’, an examination of the standard deviation associated with this estimate of bias should temper this desire. The standard deviation associated with the model’s outcome distribution

Table 3. Annual Savings Realization Rates from Monte Carlo Models

Ratio of: / *to:*	End-Use Metering / Tracking Estimate	End-Use Metering / Site Inspection	Site Inspection / Tracking Estimate
Average	78%	88%	89%
Stdev.	34%	22%	35%

suggest that the tracking estimate, while biased by only 22% on average, is susceptible to a large degree of variation. The distribution of tracking estimate bias across sites, as computed by our Monte Carlo model of annual savings, is given in Figure 3.

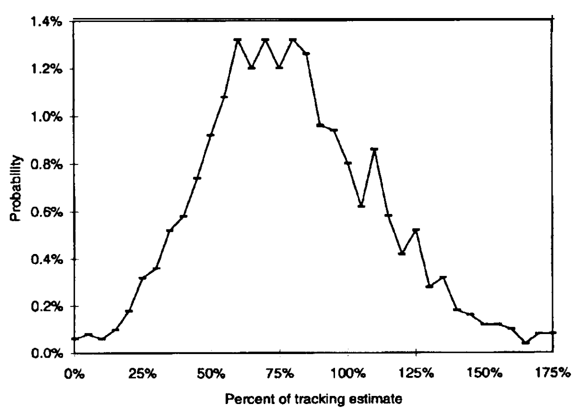

Figure 3. Distribution of Annual Savings Realization Rates

Optimally, the distribution shown in Figure 3 would be sharp and narrow (and centered near 100%), with a minimum of spread across the x-axis. But as the large standard deviation in Table 3 suggests, the distribution of realization rates is subject to a significant amount of uncertainty. If we assume the distribution is roughly normal, then only 68% of the sites have realization rates between the wide margin of 44% and 112%. As a result, applying an average realization rate gleaned from a subset of program participants to program participants at large, or to program participants of a subsequent or previous year, should be approached with caution. Even small differences between the sample population and other populations could cause relatively large differences in the average bias of the tracking estimates.

The smaller standard deviation for the outcome variable representing the difference between site inspection estimates of savings and end-use metering estimates indicates that these two methods provide more similar results.

Uncertainty Analysis

We perform uncertainty analysis by computing the rank correlation between input variables and annual energy savings for each model and examining the results. By comparing correlations between the outcome distribution and the input distributions, we can determine which input parameters contribute the lion's share of the uncertainty to the outcome distribution. If we then improve a single parameter's precision and compare correlations from different models, we can determine how valuable different evaluation techniques are in reducing the relative uncertainty of program parameters. This allows program evaluators and planners to trade off evaluation method uncertainty with method cost.

If end-use metering estimates of savings are assumed to best approximate reality, then we can interpret the rank correlations in Table 4 to mean that most of the uncertainty in tracking estimates of savings is due to misspecifications of the hours of operation parameter, and the same parameter is responsible for most of the uncertainty in site inspection estimates of savings.

An important issue for evaluation practice involves the question of whether and when to use more rigorous evaluation techniques. In this case, our analysis suggests both tracking estimates and site inspection estimates of hours of operation are subject to significant uncertainty. If data loggers, or a similar technique, provides hours of operation parameter estimates that are a significant improvement over those used in tracking estimates and site inspection estimates of savings, then augmenting tracking estimates or site inspection estimates with actual hours of operation information could result in savings estimates comparable with those obtained using end-use metering, but at a potentially lower cost.[5] Alternatively, disaggregating hours of operation by measure type or by usage characteristics may improve tracking estimates of hours of operation.

Our results would have been dramatically different if we had included NU's Energy Saver Lighting Rebate data in our uncertainty analysis: the systematic errors in NU's

Table 4. Correlation of Uncertainty in Parameters to Uncertainty in Savings

Importance of Parameter Between:	Tracking Estimate and End-Use Metering	Site Inspection and End-Use Metering	Tracking Estimate and Site Inspection
Measures per Site	.26	NA6	.27
Watts per Measure	.48	.59	.55
Hours of Operation	.82	.78	.78

tracking algorithms would have skewed the results, suggesting that most of the uncertainty in tracking database estimates of savings was due to the change in watts per measure parameter. The small sample (of three programs) which we investigate here does not enable us to determine if such systematic errors in tracking databases are a common occurrence. A benefit of the type of analysis performed here, and of the detailed site inspection and metering work performed *and reported* for these three programs in their evaluations, is that it allows tracking database accuracy to be assessed and improved.

The Effect of Measure Life Uncertainty on Lifetime Savings Estimates

The final parameter we examine in our simplified model of energy savings is the average measure lifetime. While a smattering of persistence studies exist, DSM evaluation is simply too young for any persistence studies to have followed a complete set of lighting measures through their entire estimated lifetimes. As a compromise, we examine a multi-year persistence study of lighting measures in the commercial sector. We extrapolate from this study and from tracking database estimates from NEES and NU programs to calculate optimistic and pessimistic estimates of measure lifetime.

A Long Island Lighting Company (LILCO) persistence study in 1993 examined program measures at 600 participant sites two, three, four, and five years after measure installation (Applied Energy Group 1993). Table 5 lists the percentage of measures that, as of 1993, were still functioning or had functioned for their estimated lifetime (if the utility estimate of average lifetime had already been exceeded).

Since the LILCO data represents four separate program years, we cannot make explicit time series comparisons (e.g., logically, persistence after five years cannot be higher than persistence after two years). However the data

Table 5. Results of LILCO Persistence Study

Technical persistence of operational life based on inspections in 1993

	Year Installed:			
Technology	**1988**	**1989**	**1990**	**1991**
Fl. Current Limiters	99%	100%	93%	100%
Fl. Fixture	100%	99%	100%	100%
High Int. Dis. < 200W	92%	98%	100%	100%
High Int. Dis. > 200W	100%	98%	98%	100%
Optical Reflector	100%	100%	100%	100%
CFL	99%	94%	98%	99%
34W Fl. Tubes	96%	95%	93%	86%
60W Fl. Tubes	98%	92%	78%	93%
Average	98%	97%	95%	95%
Stdev	2%	3%	7%	7%

suggests that for this sample, overall measure persistence in the first five years is probably around 95%. A study for NEES' Energy Initiative program also reports that measure persistence in the first two years is at least 95%.

Based on this short term data, we define two possible scenarios for measure lifetime.

Measure life is subject to a small level of "infant mortality" early on, but thereafter stabilizes so that average measure life is similar to tracking database estimates. This is the optimistic scenario.

Measure persistence declines rapidly after the first five years, due to remodeling, tenant changes, and premature measure failure, so that average measure life is half of tracking database estimates. This is the pessimistic scenario.

The tracking database estimates of measure lifetime for NEES' Energy Initiative and Small C&I programs, and for NU ESLR program, averaged across all program measures, are given in Table 6.

Table 6. Measure Lifetimes

Program	Average Measure Lifetime
NU ESLR	17 years
NEES EI	18
NEES Small C&I	15

For the measure lifetime parameter, two separate probability density functions were custom-specified using fractiles to represent pessimistic and optimistic scenarios of measure lifetime. The average life expectancy for the pessimistic scenario is seven years and the average life expectancy for the optimistic scenario is 15 years.

In the previous section, we analyzed the uncertainty in tracking estimates and site inspection estimates of *annual* program savings. In this section we compare the magnitude of the uncertainties in the annual savings model to those caused by our estimates of measure lifetimes in the lifetime savings model. This analysis addresses one aspect of evaluation resource allocation; the evaluation methods we've investigated so far are not used to assess equipment lifetimes. We use hypothetical distributions for the measure lifetime in combination with results from our model of annual savings to compute a probabilistic estimate of lifetime savings. Table 7 presents the results of this procedure.

The optimistic measure life estimate has only a small effect on the average ratio between end-use and tracking estimates, reducing the ratio by 9% (from 78%, as shown in Table 3, to 70%). As would be expected, the pessimistic estimate of measure life reduces the ratio of end-use metering estimate to tracking estimate by almost half. While this result is somewhat intuitive, it has profound evaluation policy implications: The large difference between the average ratios obtained when optimistic and pessimistic assumptions are made regarding measure lifetimes suggests that measure life variability may be more important than other program savings parameters. As a result, determination of which of these measure lifetime estimates is most plausible in practice should be a research priority.

Table 8 presents the uncertainty analysis results for the lifetime savings model when the two measure lifetime distributions are used. Regardless of the measure lifetime variable used, measure lifetime (as we have specified it in two hypothetical distributions) seems to be responsible for a considerable amount of the uncertainty in the lifetime savings estimate.

Conclusions

Our analysis suggests that significant uncertainty exists in 'realization rates' generated at the site level using tracking

Table 7. Lifetime Savings Realization Rates From Monte Carlo Model

Ratio of: / *to:*	Optimistic Measure Life: End-Use Metering / Tracking Estimate	Pessimistic Measure Life: End-Use Metering / Tracking Estimate
Average	70%	40%
Stdev.	38%	25%

Table 8. Correlation of Uncertainty in Input Parameters to Uncertainty in Output Variable

Parameter Importance:	**Optimistic Measure Life End-Use Metering and Tracking Estimate**	**Pessimistic Measure Life End-Use Metering and Tracking Estimate**
Measures per Site	.21	.17
Watts per Measure	.38	.40
Hours of Operation	.66	.64
Measure Lifetime	.49	.62

estimates, site inspections, and end-use metering. While an aggregate realization rate can be computed using statistical standard methods, ignoring the variability in realization rates can lead to overconfidence in the generalizability and accuracy of aggregate realization rates.

When different parameters in the annual savings equation are examined, the greatest contributor to variability in savings realization rates is the hours of operation parameter. Program evaluators interested in obtaining accurate savings estimates at the lowest possible cost could augment tracking estimates and site inspection estimates with more accurate specifications of hours of operation using data loggers, for example. Evaluators might also improve tracking database estimates of savings by using different hours of operation estimates for each measure installed, depending on measure type, building type, and the area of the building in which the measure is installed.

Few evaluations have attempted to verify the 10-20 year measure lifetime estimates used by most utilities. Due to the dramatic impact of shorter measure lifetimes on lifetime savings estimates (vis-á-vis the effects of other parameters on the lifetime savings), it would be prudent to reallocate some of those resources currently devoted to traditional impact evaluation to methods which directly assess persistence of savings in the medium and long term.

A critical understanding of evaluation method accuracy requires site-level, and even measure-level, comparisons of the parameters used to calculate savings and evaluation method results. Unfortunately, most evaluations only report aggregated results, averaged across all participants. The three reports we rely upon in this study represent what we hope will become the minimum standard for reporting evaluation results. Without sufficiently detailed evaluation data, choosing least cost, accurate evaluation methods will forever remain a matter of opinion, rather than fact.

Endnotes

1. Comparison of these methods with econometric models that utilize customer billing data requires construction of an artificial set of consumption records for a hypothetical group of customers and will not be covered in this paper.

2. Multiplying the parameter realization rates for a program yields the realization rate for the savings estimate.

3. This type of analysis would be much more powerful if end-use metering and site inspection data were readily available at the measure level. We could then estimate the uncertainty or variability in savings on a measure by measure basis. The probabilistic distributions of savings resulting from such a model would be very useful from a program planning perspective; rather than using a point estimate of savings, a probabilistic estimate could be used for each measure in demand forecasting and DSM planning and technology screening exercises.

4. Other distributions, such as a beta distribution, were found to have an improved fit, but did not affect significantly the outcome of the analysis.

5. The decision of how much to spend on evaluation should also include an analysis of the need for evaluation precision and accuracy. Some uses of evaluation information require more precise savings estimates than others; e.g., long-range demand forecast inputs vs. shared savings calculations.

6. Site inspection savings estimates and end-use metering savings estimates use the same measure per site parameter value.

References

Applied Energy Group. *1993. Persistence Study of Energy Conservation Measures Implemented in LILCO's Commercial Audit and Dollars & Sense Programs*, Long Island Lighting Company, Woodbury, NY.

Eto, J., E. Vine, L. Shown, R. Sonnenblick, and C. Payne. 1994. *The Cost and Performance of Utility Commercial Lighting Programs,* LBL-34967, Lawrence Berkeley Laboratory, Berkeley, CA.

Morgan, M.G. and M. Henrion. 1990. *Uncertainty,* Cambridge University press, New York, pp. 172-219.

Nadel, S.M. and K.M. Keating. 1991. "Engineering Estimates vs. Impact Evaluation Results: How do they Compare and Why?" *Proceedings from the 1991 Energy Program Evaluation Conference,* pp. 24-33. Chicago, IL.

RLW Analytics, Inc. and The Fleming Group. 1992. *Energy Saver Lighting Rebate: Results of the 30-Site Short Duration Monitoring Test,* C&LM Department, Northeast Utilities, Westbrook, CT.

RLW Analytics, Inc. and The Fleming Group. 1992. *Small C/I Program: Impact Evaluation Using Short-Duration Metering, New* England Electric System, Westborough, MA.

RLW Analytics, Inc. and The Fleming Group. 1992. New *England Power Service Company Energy Initiative Program: Impact Evaluation Using Short-Duration Metering, New* England Electric System, Westborough, MA.