Streamlining Impact Evaluation for Customized DSM Projects

Jeremy J. Newberger, Senior Analyst, New England Electric System

Customized energy efficiency measures implemented in the large commercial and industrial (C&I) sectors of the New England Electric System's (NEES) retail companies accounted for over 50,000,000 kWh, or 50%, of the gross C&I program energy savings achieved in 1994. The utility spent approximately 23% of the authorized incentive it collects on DSM programs on impact evaluation of these measures. In eight end-use studies done in 1995, 75 projects were evaluated—mostly by using intensive site metering techniques. The utility reviewed and streamlined its customized impact evaluation effort in an attempt to make it less costly while maintaining the quality of the results.

In this paper, several sources of discrepancy between original savings estimates and final evaluated results are described. Through a review of the site evaluation reports, a primary source of discrepancy is identified for each project. Many discrepancies would have been apparent without the use of costly, intensive site metering. Examples of specific projects within end-use categories are used to illustrate this point. Based on discrepancy classification and project type, new guidelines are developed which focus on reducing the use of short-term metering for evaluating custom projects. Compared to the NEES Companies' 1994 evaluation effort, this plan will potentially reduce the number of intensively metered sites by about 40%, with a negligible impact expected on realization rate results. An evaluation cost reduction of 16% to 20% is projected. Further cost reduction will be possible through combining two years of projects in some end-use studies.

INTRODUCTION

Background

Customized energy efficiency measures implemented in the large commercial and industrial (C&I) sectors of the New England Electric System accounted for over 50,000,000 kWh, or 50%, of the gross C&I program energy savings achieved in 1994. The utility spent approximately 23% of the authorized incentive it collects for implementing DSM on impact evaluation of these programs. In eight end-use studies conducted in 1995, 75 projects were evaluated—mostly by intensively metering electric energy of affected equipment at the site. The evaluation studies yielded a realization rate of 94% of original kWh savings estimates with a precision of +/− 9% (Wright 1995).

The utility decided to review and streamline its customized impact evaluation effort in an attempt to make it less costly by reducing site metering while maintaining the quality of the results. Although the original scope of the 1995 studies anticipated using metering at only some sites, in fact one to four weeks of metering were done at most sites because of NEES' strongly expressed preference for metering in evaluation.[1] Since short-term metering was by far the predominantly used method,[2] it was made the focus of the cost reduction effort. This year a more rigorous effort is being made to encourage alternative evaluation techniques through the development of evaluation methodology guidelines. This paper describes the guidelines and their development.

The New England Electric System operates through three retail companies: Massachusetts Electric Company, The Narragansett Electric Company in Rhode Island, and Granite State Electric Company in New Hampshire. New England Power Service Company (NEPSCo) is another subsidiary of NEES and handles many of the core tasks for all three retail companies, including impact evaluation of DSM programs. While there are differences in the programs implemented in each state, evaluation studies are usually designed to be representative and applicable for all states.

Scope

This paper focuses on the development of a plan to evaluate the 1995 custom DSM projects based on the premise that, through more thorough preparation, the use of intensive site metering as a primary evaluation tool will be reduced and, consequently, the evaluation process will be more manageable and less costly while still yielding quality results. The guidelines for screening projects and identifying situations where metering may not be required have been reviewed and accepted by evaluation contractors[3] and NEES' DSM collaborative partners, the Conservation Law Foundation. The plan is being implemented at this time. Successful implementation will be judged by the costs and results of the studies. Some studies will be completed by June 1; others will be finished by December 1.

Traditionally, the evaluation calendar lags one year behind the program year. Thus, projects implemented in 1994 were evaluated and reported on in 1995; projects implemented in 1995 will be evaluated in 1996. Therefore, ''1995 studies'' and ''1994 results'' refer to the most recently completed evaluation studies of 1994 projects. This paper presents a plan to evaluate the 1995 custom projects based on an analysis of 1994 results. While developed specifically for custom projects, the methods presented here are applicable to other evaluations that have relied heavily on site metering for results.

METHODOLOGY

Discrepancy Identification and Assignment

Eleven custom impact evaluation studies were done in 1995 on custom projects paid in 1994. Two were very limited in scope and the results of another study were not available for analysis. Projects evaluated in the eight remaining studies, encompassing 75 projects, were reviewed. The purpose of the review was to identify the reasons for the discrepancy between the tracking estimate of savings used as the basis of project approval and the evaluated results. Some projects were given more than one code, and a primary reason for the discrepancy was identified. The ''discrepancy codes'' are described in Table 1.

The assignment of the codes for the Industrial Refrigeration end-use group is illustrative of the code assignment process. The evaluation study sample for this group consisted of five projects. The original savings analyses were done by four different engineering firms and incorporated different methodologies.

- One of the projects involved modifications to refrigeration equipment at a soda bottling facility. It was one of the ten largest projects paid. The primary reason for the savings discrepancy was a change in the number of bottles of soda produced. A secondary reason was the use of incorrect feedwater temperatures in the original analysis. This project was assigned code H, since a production change was the major driver of the savings discrepancy.

- A cold storage warehouse retrofit was evaluated using a regression analysis of hourly load research meter data from before and after the retrofit. This project was assigned code C.

- Another project was a thermal shifting project at another warehouse. It was also evaluated using hourly meter data. However, the old compressors were left in place and they proved to be unreliable and incapable of doing the necessary cooling off-peak. Consequently, the thermal shifting could not occur and this project was assigned code D.

- The fourth project involved modifications at a seafood processing facility. There, the key difference was determined to be a needed adjustment to the assumed baseline heat load associated with freezing a piece of batter-coated deep-fried fish. This was not considered to be a calculation error (code B), because this microscopic difference is unlikely to have been discovered through a review of the calculations alone. The monitoring data was instrumental in pointing out the source of the difference; therefore, this project was assigned code G.

- The final project covered the new compressors and controls at a packaged meat processing facility. Everything was in place and operating. Through the use of spot metering, an in-depth interview about operating procedures, and a review of manufacturer's data, it was determined that the equipment was not operating as efficiently as predicted. This project was assigned code J.

Table 2 presents the results of the discrepancy assignment process. It shows a summary of the projects grouped by primary discrepancy category. The ten largest projects are grouped as code 10, regardless of their primary discrepancy code, because the new evaluation strategy contemplated allowing these projects to remain free of the guidelines being developed based on analysis of discrepancy types. Assignment of the special code serves to exclude them from the analysis. Table 2 shows the sums of original tracking estimates of kWh and peak kW savings and evaluated results of the same quantities and the percent difference between those quantities for each group.

Analysis of Discrepancy Types

The categorization of discrepancies supports an effort to apply these results to future evaluations. In particular, it is assumed that discrepancies attributed to reasons A through F should be evident without intensive post-installation metering. (Category A is included even though it covers projects where there was no discrepancy.) Although many discrepancies only become apparent after metering is done, if intensive site metering were limited, the discrepancy would probably be discovered through other methods. In effect, the evaluation plan for each project should be subject to certain control checks before proceeding to metering.

Category G was initially thought to be a discrepancy that could be found without metering. However, while some baseline adjustments may be readily apparent following a site visit and interview, most of the projects in this category

__Table 1.__ Discrepancy Categorization

Code	Discrepancy Description
A	OK (no discrepancy).
B	Calculation errors.
C	Load research hourly meter or billing data used to compare consumption before and after project installation. Post-installation data would not have been available to the analyst developing the original savings estimate. If the discrepancy is relatively small, it may be attributed to the different method alone. If the discrepancy is large, it would probably be assigned to another group code.
D	Inoperable equipment.
E	Misapplication of technology; shouldn't be expected to save energy. This includes such things as placing liquid pressure amplification pumps on water cooled condensers.
F	Equipment changes from application, including count.
G	Adjustments to temperatures, production levels, operating conditions, and other baseline assumptions used in the original analysis and/or other methodological differences. Evaluators were instructed to use original analysts' methods and verify baseline quantities where possible. In some cases, adjustments were necessary.
H	Hours/load/operation changes.
I	Sub-optimal control of operating equipment.
J	Equipment performance different from predicted.

were complex and probably needed metering to identify the discrepancy. The frozen fish processing project is a case in point. Therefore, categories G, I, and J represent projects where the discrepancies would only be apparent following metering and for which metering ought to be a prescribed activity.

Category H, a change in loads or hours, may or may not be revealed through intensive metering. Loads discrepancies may only be apparent through metering, while significant hours changes may be apparent from discussion with facility operators. Therefore, an evaluation method without intensive metering may be appropriate for some of these sites as well. Furthermore, variations in loads and hours may occur randomly over the life of a project (where it is not explicitly stated that a customer is expanding or contracting operations). For these projects, it would be appropriate to allow a band of tolerance around the original estimate of loads or hours.

Guidelines for Future Custom Project Impact Evaluation

Based on the number of projects in the 1995 studies which showed discrepancies attributable to categories A through F, a percentage of projects in each end-use group which could be evaluated with intensive site metering was developed. For example, the percentage of projects with discrepancy codes A through F in the process cooling end-use was 75%. Under a rationing plan, the evaluator would have been allowed to choose to use short-term metering at only one of four sites in the 1996 study of 1995 paid projects. The remaining projects would have to be evaluated using other methods.

This approach was rejected as being incompatible with the variety of projects encountered through custom applications; they are, by definition, unique and subject to change from year to year. It is unrealistic to expect that reasons for discrepancies in each end-use group would follow the same pattern from one year to the next. Furthermore, proscribing the evaluation activities of evaluation contractors by rationing the number of sites at which metering is permitted would rob them of their freedom in choosing appropriate evaluation methods and possibly compromise results.

Consequently, another approach was developed. Guidelines were written for evaluators to *consider* using evaluation techniques other than site metering and incorporated into

Table 2: Discrepancy Code Analysis Summary

Discrepancy Code	Count	Pre-eval. kWh	Post-eval. kWh	% diff. from pre-eval.	Pre-eval. kW	Post-eval kW	% diff. from pre-eval
10	10	17,592,347	18,335,941	4%	2,936	3,485	19%
A	2	244,577	244,278	− 0%	27	28	4%
B	11	1,524,279	815,357	− 47%	264	184	− 30%
C	5	369,359	374,107	1%	45	49	9%
D	1	41,240	41,240	0%	86	0	− 100%
E	4	191,759	24,670	− 87%	26	5	− 80%
F	1	405,899	11,688	− 97%	168	102	− 39%
G	6	1,752,370	1,491,064	− 15%	778	529	− 32%
H	17	4,209,667	3,869,613	− 8%	742	802	8%
I	8	919,534	482,269	− 48%	185	89	− 52%
J	10	2,147,700	1,895,965	− 12%	407	400	− 2%
TOTAL	75	29,398,731	27,586,192	− 6%	5,665	5,674	0%

the scope of work for the impact evaluation studies to be performed in 1996. Although mere consideration does not ensure the certainty of evaluation streamlining as rationing would have, it is expected to offer significant savings, while maintaining the quality of results.

The new scope of work for custom project evaluation instructs evaluators to thoroughly review the application material for each project in the study. Based upon a review of the material, along with an optional preliminary scoping conversation with facility operators, evaluators are to develop an appropriate evaluation plan for the project, and submit it to NEPSCo for review. Projects greater than 1,000,000 gross kWh may be evaluated using all appropriate techniques, unrestricted by guidelines. For the remaining projects, evaluators are encouraged to investigate ways to reduce, where possible, the use of intensive site metering as a primary evaluation tool in the evaluation studies.

As the application material is reviewed, and plans formed, the following criteria are to be considered by the evaluators. They serve as ‘‘off-ramps’’ from an evaluation track which otherwise would have used intensive site metering. Each of these criteria focuses on identifying discrepancy types which are observable without metering.

To short circuit studies with major calculation errors (code B):

- If a substantial math error is found during application review, contact the study manager to discuss the future focus of the evaluation. If much of the savings has been discounted through such an error, there is diminished value to long term metering. On the other hand, it may be necessary to continue evaluating due to magnitude and uncertainty of the remaining savings.

To steer evaluators toward use of hourly meter or billing data or other post-installation metering data (code C):

- For a retrofit program project, if the energy savings represent 10% or more of the facility usage and has a distinct impact on the facility load shape, consider using before- and after-retrofit whole building meter hourly data in the evaluation.

- For a new construction/renovation/failed equipment replacement program project, if the project represents a substantial portion of the load in a renovated or new facility and has a distinct impact on the facility load shape, consider using whole building meter hourly data in the evaluation.

- If metered data is available as part of the material provided about each project or from the site contact, structure the evaluation using it. Determine if it is valid and representative through interviewing the party who collected the data and use it wherever appropriate. Replication of metered data considered to be valid is not desirable.

To cut short studies that may waste effort evaluating projects that don't work (codes D and E):

- If based on review of the material, it cannot be determined how the project saves energy, focus the evaluation first on answering this fundamental question. If in fact the project is found to have no substantial impact on energy consumption, consider truncating the study.

- If once the site is visited, the equipment is found to be inoperable or installed incorrectly in such a way as to make it obvious that there is little or no impact on energy use, the study should be truncated after notifying the study manager.

To determine the impact of equipment changes (code F):

- If the equipment is significantly different from what was specified in the application (model, size, efficiency, or count), this may substantially affect savings. Focus the evaluation on the key parameters. It may still be necessary to complete the evaluation due to remaining uncertainties about equipment operation.

To assess the relative importance of hours or loads changes before metering (code H)

- Before doing extensive run-time metering, determine through interviews with facility personnel whether run-time is expected to deviate appreciably from application estimates.

- In cases where the difference in hours of operation or loads is small (within 10% of the assumption used in the tracking estimate of savings) and where the variation is considered to be random and is likely to be different using the same measurement protocol at another time, use the original estimate of hours or loads. Using the measured quantity implies a level of precision which cannot be claimed from observation over a relatively short period of time. Variances which are observed in the course of evaluation activities which meet these conditions will be documented, but not used in the calculation of savings.[4]

RESULTS

Expected Reduction in the Number of Intensively Metered Projects

The data in Table 2 suggest that using methodology screening criteria will lead to between 37% and 63% of the projects outside the top ten being evaluated without intensive metering at the site. Of the 65 projects studied, 24 (37%) showed discrepancies which could have been observed without metering. An identical number fell into categories that needed metering to identify the discrepancy. The remaining 17 projects (26%) fell into category H, a change in hours or loads, which may or may not be revealed through intensive metering. Based on this analysis, we expect evaluators will be able to limit intensive site metering to no more than 50% (the midpoint of the 37% to 63% range) of the projects in the overall sample. Overall, the metering percentage will be about 60% of the projects, when the largest ten projects are added. The evaluation contractor will have a choice as to which projects should be metered.

Minimal Risk of Inaccurate Realization Rates or Relative Precision

While non-metering evaluation techniques can identify the source of discrepancy, there remains some uncertainty whether the process will yield results that are identical to an overall evaluation strategy which states ''go and meter all sites''. Theoretically, the results—both realization rate and precision—should be the same as with metering of all projects. For the approximately 60% of projects which will still be metered they will be identical. If a project doesn't have to be metered but is, the results should be more than adequate (although evaluation cost savings would be forfeited).

If a project is among the 40% that are not metered, there is a chance a numerically incorrect result, as measured against the standard of a metered result, will be determined. The incorrect result would be due to either being constrained to use an inappropriate method, or using a method that fails to provide adequate data to calculate correct savings. However, the screening criteria are well suited for identifying discrepancies of large magnitudes. As seen in Table 2, the overall discrepancies for categories B, D, E, and F were very large, ranging from −47% to −100% (looking at the capacity discrepancy of the industrial refrigeration load shifting project in category D).[5] The magnitude of overall discrepancies in the code groups which use metering is smaller, ranging from −12% to −48%. Metering sites where larger discrepancies may be found would not be likely to change results much and is not warranted.

The logical exploration of the risk gave us the confidence to proceed with this plan. No control study or simulation of results is planned to compare metering and non-metering results for the same sites. The quality of the results will be judged by comparison with the prior year's results and, ultimately, the acceptance of evaluation results when filed. Interim realization rate results for the 1996 studies will begin to be available in June, 1996. Interim results for other studies will be available in December, 1996. Because many studies are being performed on a two-year cycle, realization rate and precision results will not be final until final results are in on all 1996 projects, in mid-1997.

Evaluation Cost Reduction

While intensive site metering is the most expensive evaluation technique, even for those sites where metering is not to be used some evaluation technique will be employed. In other words, reducing metering at 40% of the sites will not lead to a 40% cost reduction for evaluation studies. We estimated that costs will be reduced by 40 to 50% for the sites where metering is not being performed, for an approximate savings of 16 % to 20%. The cost savings will be analyzed by comparing the final cost per site for the 1995 and 1996 evaluation studies for each end-use group.

As of June, 1996, eleven end-use studies[6] are being conducted using the guidelines for screening projects and identifying non-metering evaluation methodologies. The budgeted costs per site for these studies are shown in Table 3, compared where applicable to final 1995 study costs per site. While cost reductions have been realized in some end-use groups as a result of the guidelines, some groups have seen study costs per site increase, owing to the magnitude of the sample projects and the comfort level of evaluators with the new guideline process. Overall, the evaluation budget for 1996 is roughly the same as 1995 while average budgeted costs per site have decreased 7%. Total evaluation costs for the eight groups studied in both 1995 and 1996 have decreased 9%. Actual costs for 1996 may decrease slightly as evaluations proceed and some of the control checks in the guidelines are triggered.

Additional cost savings will be realized because NEPSCo will be performing some studies in 1996 which will be designed to be used in reporting the results for custom projects paid in both 1995 and 1996. Some of these studies will be in end-uses whose savings are often cooling season dependent. Others are in areas where results have not shown much change from year to year or where participation is low. Consequently, there will be only one study for the two-year period in seven end-use groups. Over the two-year period, this will lead to significant cost savings even though the evaluation period for some of the weather-dependent end-use groups will be somewhat protracted, with associated costs.

CONCLUSIONS

The use of guidelines for screening projects and identifying non-metering evaluation methodologies has led to a preliminary reduction in cost per site of 7%, compared to predicted reductions of 16% to 20%. While the guidelines have influenced cost reductions in some areas, costs are also influenced by the size of the projects being evaluated and the comfort level of the evaluator with the guideline process. Overall evaluation costs for same-group studies have decreased. However, the total evaluation budget is primarily subject to decisions about which groups to study and when. For example, the end-use group with the highest cost study in 1995 is not being studied in 1996, due to much lower participation. The highest cost study in 1996 is a new study in the evaluation line-up. The combination of study years for some end-use groups will lead to a significant reduction in evaluation costs in 1997.

Because the types of discrepancies which are intended to be captured by the screening criteria are usually of larger magnitude than those discrepancies identified through intensive site metering, we do not think there will be an appreciable impact on either realization rate or relative precision results. Preliminary results will begin to be available in the summer of 1996. Final results will be available in mid-1997.

It is encouraging to note that the guidelines have the support of NEES' evaluation contractors as well as consultants to the Conservation Law Foundation who monitor the impact evaluation process as collaborative partners in NEES' demand-side management program implementation. The NEES Companies continue to work with these other parties in the evaluation process to streamline it and achieve cost reduction while maintaining the high quality of results.

ACKNOWLEDGMENTS

The author acknowledges the contributions of Dottie Conant, Ann Hatcher, and David Jacobson at NEPSCo for their constructive review of the evaluation plan as it was developed.

ENDNOTES

1. NEES allowed evaluators much latitude in evaluation. Evaluators were asked to try to use the calculation

Table 3. Impact Evaluation Study Cost Per Site

End Use Group	Study	1996 Cost/ Project	1995 Cost/ Project	1996/1995	1997 (a)
1	Compressed Air	$4,790	$4,046	118%	No
2	Motors	Not studied	Not studied	N/A	No
3	Process Cooling	$11,607 (b)	$6,667	174%	Yes
4	Process & Other	$3,942	$4,231	93%	Yes
5	Commercial Refrigeration	$2,647	$2,680	99%	Yes
6	Industrial Refrigeration	$10,728	$17,600	61%	No
7a	Variable Speed Drives(Custom)	$4,757 (c)	N/A	N/A	No
7b	Variable Speed Drives (Prescriptive)	$524 (c)	$1,255	42%	No
8	EMS (Custom & Prescriptive)	Not studied	$16,000 (d)	N/A	Yes
9	HVAC	$4,675	$6,846	68%	No
10a	Comprehensive Design Approach	$15,000	$18,100	83%	No
10b	Chiller Initiative	$24,121 (d)	Not studied	N/A	No
11	Lighting	$1,374	Not studied	N/A	No
	Average for all studies:	$4,859	$5,222	93%	
	Total costs for all studies	$709,381	$704,961	101%	
	Total costs for 8 same studies (included VSDs)	$482,733	$528,961	91%	

Note: All study groups for custom projects unless otherwise indicated.
(a) Indicates whether study planned for this end use group in 1997.
(b) Evaluated project savings 3 times greater than in 1995 study.
(c) Custom VSDs studied separately for first time. 1995 study used same approach (telephone survey for data collection) for custom and prescriptive.
(d) Highest cost study this year

method used in the original savings estimate. However, metering data was allowed to supersede other sources of information, thus the emphasis on metering. Evaluators were allowed to choose the length of the metering period for each project. Because of the relatively short time frame for metering, evaluators were always required to justify that conditions which existed during the metering period—such as production or weather—were representative of typical operating conditions. For weather-dependent measures, metering period results were calibrated to the weather-sensitive modeled performance for the same period found in the original estimate of savings.

2. Other techniques used included analysis of whole building load research data, building simulation models, on/off data run-time loggers, review of production logs, and spot metering. No attempt was made to isolate the costs of each technique.

3. NEES hires independent outside engineering firms to perform the impact evaluations for custom energy efficiency projects. Firms were selected mostly through an RFP process in 1995 and did between one and three studies each. Firms were screened for conflict of interests with the sites and end-use groups they were chosen to evaluate prior to the commencement of work. Most of the evaluation contractors used in 1995 were awarded the identical studies in 1996.

4. For example, if the tracking estimate of savings said the equipment would operate 5000 hours and measurement show it was operating at an annual rate of 4950 hours or that throughput was predicted at 2000 lbs/hr but turned out to be 2100. The evaluator should consider these differences and judge whether they are truly representative of the status quo operation. If not, the measurement should be noted in the report but not used. If yes (for example, the throughput profile was flat during the monitoring period), the evaluator has firm reason to believe that the observed level better represents the status quo and the measurement should be used.

5. Category C, use of hourly billing meter data, is excluded here. While the use of this data may lead to results which are different from predictions, it is merely considered to be different, not incorrect.

6. Of the eleven end-use categories in the 1995 evaluation cycle, two were combined, and two were split into subgroups, leading to a maximum of 13 study groups. Of these, two groups had very low participation in 1995 and are not being subject to evaluation study in 1996.

REFERENCES

Wright, R.L. 1995. Analysis of Custom Retrofit Measures. Prepared for New England Electric System, October 23.