NORTHEAST UTILITIES' APPROACH TO C&LM PROGRAM IMPACT EVALUATION

John Amalfi
Northeast Utilities
Rocky Hill, Connecticut

Roger L. Wright
RLW Analytics, Inc.
Sonoma, California

Introduction

Northeast Utilities (NU) is New England's largest utility, serving more than 1.2 million residential, commercial, and industrial customers in Connecticut and Massachusetts. It is charged with providing dependable and reasonably-priced energy to meet the long-term needs of the region. For more than a decade, NU and its regulators have recognized the important benefits of conservation and load management (C&LM) in helping to meet this challenge.

Beginning with the October 1980 DPUC order in Connecticut to submit a "comprehensive and cohesive conservation program," C&LM activities at Northeast Utilities grew steadily through the first half of the decade. In 1986 the *Energy Alliance* umbrella of C&LM programs and services was formally introduced. In 1988 the collaborative planning process between the Company and non-utility parties was kicked off in both Connecticut and Massachusetts, resulting in significant "ramping up" of program implementation and expenditures. In 1991, for example, NU has earmarked more than $75 million for C&LM systemwide; in 1990, the Company achieved more than 2.5 million MWh in lifetime savings from a comprehensive set of programs built on national industry expertise.

Today, because C&LM is a very real part of NU's commitment to its customers, considerable effort will be spent to evaluate the on-going effectiveness of individual C&LM programs. The Company filed its most recent Interim Evaluation Plan in Massachusetts in March 1991 and circulated the plan to collaborative parties in Connecticut. As documented in this Plan, NU's evaluation efforts are intended to meet several important objectives, including:

- Provide timely feedback of information to internal and external decision makers in order to improve C&LM program design, implementation, and evaluation.
- Provide information essential to quantifying C&LM's contribution to integrated resource planning and cost-effectiveness testing.

To achieve these objectives, NU's C&LM evaluation activity is concentrated in three inter-related areas: process evaluation, impact evaluation, and tracking systems. Together, these activities provide the data relevant to understanding program performance, energy and capacity savings, and the costs to achieve these savings.

The focus of this paper is NU's approach to impact evaluation—how we quantify a C&LM program's impact on energy and demand consumption. We will discuss techniques the Company is currently using to fulfill different information needs, and how we integrate various sources of information to achieve the most useful results from our evaluation expenditures.

Impact Evaluation: An Overview

NU has identified specific impact evaluation goals to guide its efforts, including:

- Develop reliable, unbiased estimates of program energy savings and demand reductions.
- Provide information on the technological capabilities of various energy conservation measures (ECMs).
- Improve the accuracy of methodologies and techniques used to estimate impacts.
- Ensure a consistent framework for the assessment of C\&LM program impacts.

NU believes that balanced, insightful evaluation should integrate information from various sources of data and methods of analysis—and that, as in other tasks, picking the right tool for the right job is of paramount importance.

Three impact evaluation techniques which NU is currently using are discussed below: engineering estimation, billing analysis, and end-use metering. Advantages and disadvantages of these techniques are summarized, and strategies for improving accuracy are illustrated with working examples. Of particular interest is the discussion of the Engineering Calibration Approach (ECA), a technique being developed by the Company to integrate the results of various methods into a single consistent estimate of program impact.

Engineering Estimation

Engineering estimates of impact are developed using a variety of approaches tailored to individual C&LM programs. Estimation procedures differ between prescriptive and customized programs, but in general rely on estimates of measure life, energy consumption differences between old and new measures, and operating hours to arrive at estimated program impact. Specific estimation methodologies are based on general engineering knowledge; customer surveys, including site visits; and NU's research through billing analysis and load research.

The advantages of using engineering estimates are clear. Data are instantly retrievable through program tracking systems, and calculations are quick, inexpensive, and accurate for simple applications. Disadvantages of this technique include the difficulty in obtaining good information on product replacements, actual operating hours, and interactive effects.

Given current technology, engineering estimation is a prudent tool for assessing the impacts of certain conservation measures such as residential lighting. NU's Lighting Catalog Program illustrates an appropriate application of this technique.

The Lighting Catalog Program targets efficiency improvements in one end use of the residential sector, and of course "before" and "after" wattage reductions are straightforward. The Company is refining the engineering estimates now in use through customer surveys and focus groups. The data obtained will help us improve assumptions about measures actually installed, daily operating hours, and measure life. In turn, the refined estimates will upgrade our estimation of impacts across the spectrum of residential programs, since lighting is a growing source of savings in this sector.

Another application of engineering estimation can be illustrated with savings associated with residential electric water heaters. NU performed an analysis of nearly 14,000 wrapped water heaters in its service territory, and, based on ASHRAE engineering estimates, kWh savings were calculated according to 18 categories of participants. The characteristics used to define participant categories were: (1) the location of the water heater (*i.e.*, whether in conditioned or unconditioned space); (2) the temperature before and after the turndown; and (3) the size of the tank (with three ranges of sizes represented).

Based on these inputs and periodic updating of assumptions, the average annual savings per participant was determined to be 654 kWh. We believe these estimates can be enhanced with the results of further impact analysis, and that billing analysis may be an appropriate tool in this instance.

Billing Analysis

Billing analysis utilizes energy consumption data from customers to estimate impacts from efficiency improvement installations. It is generally used to estimate the impact of programs with large numbers of participants for several reasons. First, since little or no new data collection is required, evaluation costs can be kept down—it is a mid-priced approach, in between engineering estimation and end-use metering. Secondly, because of the lower cost, larger samples can be analyzed, thereby enhancing the accuracy of estimates and, under ideal circumstances, even revealing impacts that otherwise might be lost in the "noise" of customer behavior, weather, and other factors. Moreover, contrasting large non-participant samples may reveal the effect of improvements being made outside the program, and thus are sometimes used to help estimate free riders. Some disadvantages of billing analysis for use in impact evaluation are that billing data measure total consumption rather than consumption at the ECM level, and provide only monthly, rather than daily or hourly, consumption levels.

Billing history databases provide extensive, in-house sources of customer behavior information. Typically, billing data relate to total building energy uses and are analyzed either by aggregating an account's monthly meter readings to an annual period, or by using the disaggregated monthly readings as time-series measurements of energy use. When annualizing meter readings on an account-by-account basis, programmatic effects are studied by employing cross-sectional, quasi-experimental research designs. In studies of this kind, annual energy consumption for program participants is compared either to annual consumption prior to participation in a program, to annual consumption of a matching control group of non-participants, or to both. The critical assumption underlying these "simple" research designs is that, but for the program, all factors related to energy consumption were, and remain, equal between participant and non-participant groups during pre- and post-program years.

Although the "all things being equal" assumption is useful as a starting point for a cross-sectional research design, program evaluations that strive for accuracy require that certain factors be explicitly taken into consideration in the energy use analysis. For example, weather typically affects energy use and varies from year to year. Thus, all energy-related factors are not equal from year to year—annual consumption must be normalized to average, long-run temperatures in order for the inter-year comparisons to be reliable.

Many other factors complicate comparisons of annual energy use among years and among study groups. Once the assumption of "all things being equal" is relaxed, annual energy use must be examined by controlling for key variables that differentially influence energy consumption, such as building size, number of occupants, and hours of use of major equipment. Practically speaking, this can be done sequentially or, if three or more key variables are involved, by controlling for their impacts jointly using a multivariate statistical model.

In a typical regression model of this kind, annual energy use, or the change in annual energy use, is specified to be a function of program participation, weather, equipment use, and customer characteristics. However, this level of analysis requires supplemental customer data beyond that normally available in billing account databases. These data may exist in utility databases such as those created for audit programs or appliance saturation surveys. Often, a primary data collection effort is required via mail or telephone survey.

A more complex method of analyzing billing histories involves using periodic meter readings in an elaborately specified time-series cross-sectional research design. This method, known as "conditional demand" modeling, offers the possibility of using billing histories not only to study program-related changes in total building use, but to disaggregate total use into its major end-use components. It requires large sample sizes and detailed data on customer characteristics and appliance holdings.

In addition, load metering data and engineering estimates can be introduced into these models in a variety of ways to improve the estimation procedure. The information contained in these data is used to constrain key parameters, forecast unobserved values, or otherwise adjust a model's estimates, given prior knowledge of energy use relationships. Finally, to supplement conditional demand models, discrete choice models can be developed to control for self-selection bias or to control for situations that involve multiple discrete decision levels. As these models require the same type of data that are used in the conditional demand models, extra data collection is unnecessary. The products of the self-selection for "nested" models are correction factors that are inserted into the conditional demand models to improve the overall estimates of program-related savings.

As part of its initial efforts in 1990 to incorporate billing analysis into NU's evaluation efforts, a billing analysis was conducted for the EnergyCHECK program, which provides audits and measure installation for small commercial customers. The sample included 100 participants and a control group of 100 nonparticipants matched for geographic location, building type, building square footage, and annual energy usage.

The general strategy for the participant's billing analysis followed a classical pre- and post-treatment experimental design, comparing annual building energy usage before and after the retrofit. To strengthen this research design, the participant's usage was compared to the control group's usage. In addition, a great deal of data preparation preceded the analysis, consisting of screening and "cleaning" the data, sorting dates and establishing pre- and post- cutoff periods, and statistically annualizing and weather-normalizing the billing histories.

To further refine this evaluation's estimate of energy savings, site visits and surveys with all of the sample's participants and non-participants were conducted. The Company's evaluation contractors verified all measures actually installed, and sought updated information on factors affecting energy usage. After annualizing and weather-normalizing customer billing histories, mean adjusted kWh usage for program participants was found to decline by 1,357 kWh/year, while control group usage increased by 3,927 kWh/year. "Net" program savings, or the change in consumption attributable to the EnergyCHECK program, was calculated by combining the annual changes of the two groups. This resulted in average net energy savings of 5,284 kWh/year per building.

To control for building size, a second energy use analysis was performed. The data indicated that participants and non-participants in the pre-implementation period tended to use electricity with about the same level of intensity. However, in the post-implementation period, clear differences in energy intensity were found between the two groups (*i.e.*, participants' use decreased by an average of 0.25 kWh/ft^2/year, while control group use increased by 0.98 kWh/ft^2/year). This difference was highly statistically significant. The net program impact, calculated by combining the annual changes for the two groups, resulted in average savings per building of 1.23 kWh/ft^2/year, suggesting that the EnergyCHECK program was responsible for a 9.66% reduction in energy intensity for the participants from the pre- to the post-implementation period.

To avoid double counting the energy savings attributable to NU's commercial lighting rebate program, estimated lighting rebate savings were factored out of the estimates of energy savings for the EnergyCHECK program. The analysis found a "net-of-net" estimate of program-related energy savings of 3,012 kWh/year or an average annual reduction in energy consumption of 4.25% per building. In terms of energy use intensity, the analysis found net-of-net savings of 70 kWh/ft^2/year, or an average annual reduction of 5.5% per building.

Billing analysis is an integral part of NU's efforts to refine engineering estimates of savings, to develop statistically adjusted engineering estimates, and to add precision to the engineering calibration approach described below.

End-use Metering

End-use metering is generally considered to provide the most accurate approach to measuring the impact of a C&LM project. NU's approach is to monitor the hourly energy and operating conditions of the end uses affected by each energy conservation measure on a before/after basis. The principal disadvantage of end-use metering is the total cost of collecting and analyzing the data. The increased precision and level of detail of actual time-of-use and end-use consumption does not come cheap.

End-use metering can be used most efficiently by tailoring the monitoring plan to the measurement objectives and the nature of the ECMs. The Company is employing a combination of short-duration and longer-duration metering. The principal advantage of short-duration end-use metering is that site-specific results can be developed rather quickly for appropriate ECMs such as nonseasonal lighting and motor retrofits. However, even with lighting and motors, short-duration end-use metering is not suitable for measuring theinteraction of these measures with HVAC. Therefore, NU is also using longer-duration end-use metering where it is required; *e.g.*, to measure the effect of major HVAC retrofits, to learn more about the interaction of lighting with HVAC, and to study the persistence of savings.

NU is using short-duration, end-use metering and on-site surveys in the Energy Saver Lighting Rebate (ESLR) Program. The results of the metering and on-site surveys are being compared to the tracking information and engineering calculations for each of the sample projects. By investigating the discrepancies by project and looking for common factors, the Company has identified specific improvements in the rebate form and accompanying instructions, in the engineering calculation assumptions, and in the program delivery.

The Engineering Calibration Approach

The Company is also maximizing the benefits of end-use metering through the Engineering Calibration Approach (ECA). ECA is a technique to measure and understand the total DSM program impact from a small sample of projects. The impacts in the sampled projects are determined by end-use metering or other suitable evaluation methods. The sample results are extrapolated to the program by relating the impacts measured in the sample to engineering estimates developed for all program participants or for a larger sample of participants.

ECA integrates:

- *End-use metering*, which is limited to a relatively small sample but provides a benchmark to true up the engineering estimates.
- *Engineering estimates from tracking information*, which link the end-use-metered sample to the target population.
- *On-site surveys, billing data, statistically adjusted engineering (SAE) estimates, and so on*, which strengthen the results.

The ECA approach usually employs:

- *Statistical sampling*—to minimize selection bias and provide measurable precision,
- *Stratification by estimated impact*—to control the size of projects in the sample,

- *Ratio or regression estimation based on engineering estimates*—to control project-to-project variability, and

- ***Double sampling***—to integrate the results of the on-site surveys and billing analysis.

An important advantage of statistical sampling is that standard methods are available for evaluating the precision of the estimates of population characteristics developed from the sample data. But while statistical precision can always be improved by increasing the sample size, a four-fold increase in the sample is usually required to improve the precision by a factor of two. Typically, large sample sizes are required to achieve reliable results by increasing the sample size alone. This is especially problematic for end-use metering, where costs can quickly become prohibitive. Consequently, many prior end-use metering studies have not been designed to yield statistically reliable results, especially in the diverse commercial sector. NU's Energy Saver Lighting Rebate Program (ESLR) provides a good example of how the Company is addressing this issue using the ECA technique.

An ECA Application

ESLR provides rebates to customers who install efficient lighting to replace existing systems. At the time of our analysis, the available data from the ESLR tracking system for 1990 showed more than 4,700 rebates. The total annual savings produced by these rebates exceeded 90,000 MWh, calculated from the rebate applications and engineering assumptions. Northeast Utilities is implementing a study to determine the accuracy of its estimates of savings.

In the ESLR study, the Company is:

- working with a sample of 30 ESLR projects,

- carrying out a detailed on-site space and lighting survey before and after each sample retrofit,

- end-use metering a portion of the affected lighting in each project for one or more weeks before and after the retrofit, and

- using ECA to extrapolate the sample results to the program.

At the present time, complete data are available for only 20 projects. The 20 sample sites are remarkably diverse, including a doctor's office, a machine shop, a grocery store, and several restaurants. Hours of operation range from 14 to 142 per week, and building sizes are from 2,000 to 140,000 square feet. The energy conservation measures are also diverse: large numbers of T-8 fluorescent systems and electronic ballast replacements, and a few compact fluorescents, exit signs, daylight sensors, and reflectors. The lighting load profiles also reflect the diversity of the sample. Figures 1 and 2 show the monitored weekday profiles before and after the retrofit for two sites.

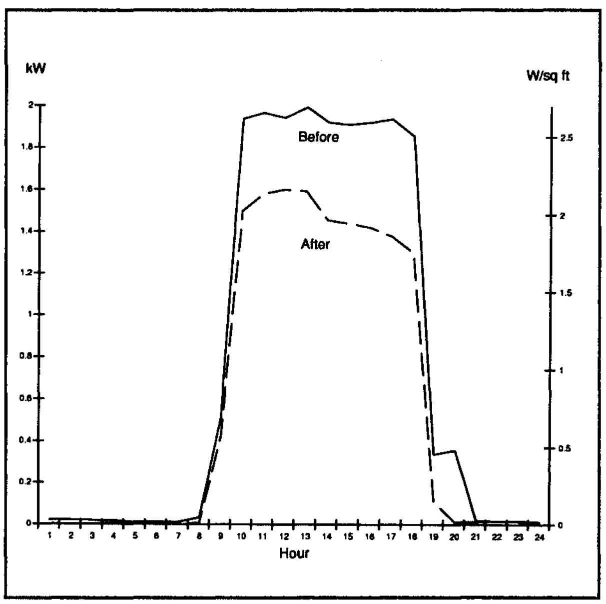

Figure 1. Site 709: An Office

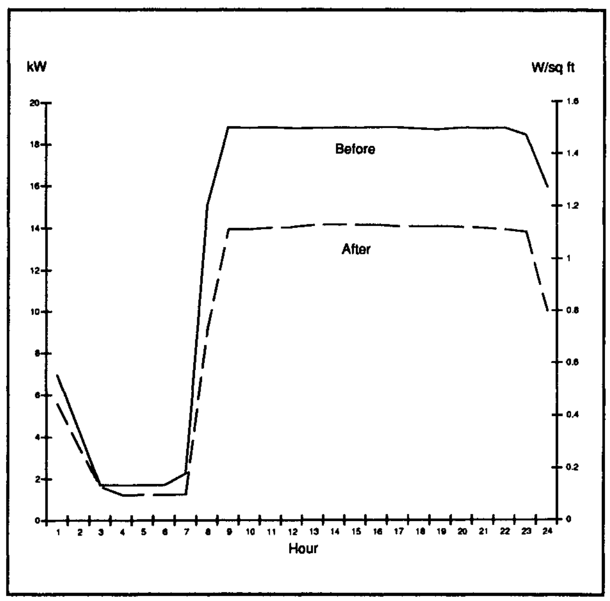

Figure 2. Site 716: A Grocery Store

Table 1. Ratio Estimation

Method	Average Savings (MWh)	Relative Precision
Average Savings:		
All Projects in the Population (N = 4764)	19.093	
Supporting Sample (N = 20)	18.365[a] 10.642[b]	±9.3%
Combining the Information:		
Estimated Total Savings		±9.3%
4,764 • 19.093 • 10.642/18.365 = 52,710 MWh Confidence Interval: 47,786 - 57,634 MWh		

[a]From tracking information.

[b]From end-use metering.

Although the study also is designed to measure demand impacts, our discussion focuses on annual energy savings. Table 1 illustrates how ratio estimation is used to link the end-use metering to the tracking information. At the time of analysis, the 1990 tracking system contained 4,764 projects—the target population. The engineering estimate of total annual savings was found to be 90,957 MWh using population tracking information. From this information, the engineering estimate of average savings can be calculated as 90,957/4,764 = 19.093 MWh per project. It should be noted that the 90,957 MWh does not take into account free riders or the rebates that were manually entered into the tracking database.

The key statistic is the ratio of metered savings to tracking system savings: 10.642/18.365 = 0.5795. In other words, the end-use metering indicates that the actual savings are about 42% smaller than the engineering estimates calculated from the tracking information for the same projects.

Ratio estimation provides a simple correction for the bias in the engineering estimates. The ratio estimate is obtained by multiplying the estimate of total annual savings calculated from the tracking information by the ratio 0.5795. For example, using gross 1990 savings available at the time of analysis, we get an unbiased estimate of total savings: 90,957 x 0.5795 = 52,710 MWh.

Table 1 also shows that the ratio estimator has achieved a statistical precision of about ±9% with a sample of only 20 projects. The statistical precision of the ratio estimate has been determined by assuming that the sample of 20 projects is a stratified random sample from the 1990 target population. The corresponding confidence interval for the total savings is from 47,786 to 57,634 MWh. This interval estimate has been calculated at the 90% level of confidence using the t-distribution with 17 degrees of freedom.

How Much End-use Metering Do You Need?

Can a sample of only 20 projects actually give reliable results?

Yes, under the right circumstances.

Consider an idealized example. Suppose that the scatter plot of savings in the target population appeared as in Figure 3. The vertical axis shows the savings that would be determined for each project by end-use metering, and the horizontal axis shows the savings calculated from the tracking information for the project. What would be the precision of the ratio estimate in this case? It is easy to see that the estimate would be perfectly even from a sample of just one project.

In practice, the relationship may not be as strong as shown in Figure 3. But one can easily see the strength of the relationship by looking at a scatter plot of the sample data. Figure 4 shows the scatter plot of kWh savings for the 20 sites in the sample. It indicates that there is a fairly strong relationship between the savings calculated from the end-use metering and the savings calculated from the information in the tracking system.

The solid line in Figure 4 corresponds to the ratio of 0.5795 previously discussed. The dashed lines in Figure 4 display a one-standard-deviation range for the expected savings of individual projects. This one-standard-deviation interval is at ±23% of the expected savings. The statistic 23% is the key measure of the strength of the relationship. This is called the *error ratio* of the relationship.

Under reasonable assumptions, there is a simple relationship between the error ratio C, the planned sample size n, and the expected relative precision R of the ratio estimate of the total savings of all projects in the population:

$$R = Z\frac{C}{\sqrt{n}}.$$

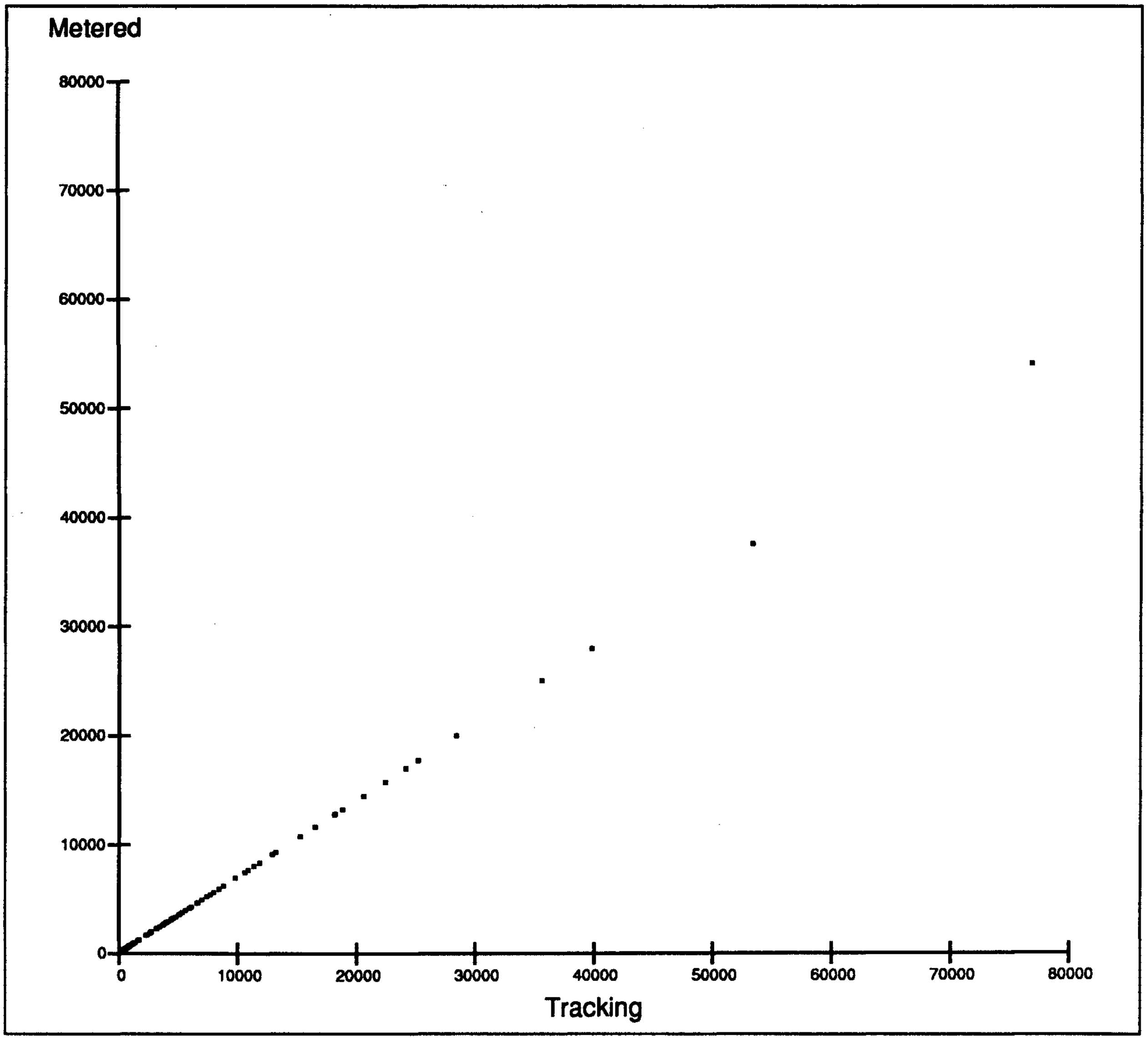

Figure 3. An Ideal Case

Here z is the normal coefficient associated with the confidence level, *e.g.*, 1.645 for 90% confidence. This result assumes the model suggested by Figure 4, a model-based sampling plan appropriately stratified by estimated savings, and ratio estimation.

The preceding equation provides a simple way of choosing the sample size when ratio estimation is used with a suitably stratified sampling plan. The sample size required for a set relative precision is:

$$n = (z \frac{C}{R})^2$$

For example, the preliminary results indicate that total savings can be estimated with a relative precision of ±10% with a sample of about $[1.761 \times (0.23 / 0.1)]^2 = 17$ end-use-metered projects. (We have used the t-statistic 1.761 here rather than the z-value of 1.645 since the sample size is so small.)

Of course, the initial sample provides only a preliminary estimate of the true error ratio in the population. If the true error ratio is actually larger, a larger sample would be required or a lower level of precision would be achieved. For example, if the error ratio is 60%, an end-use-metered sample of about 100 projects would be required for ±10% precision. In this case, a sample of 25

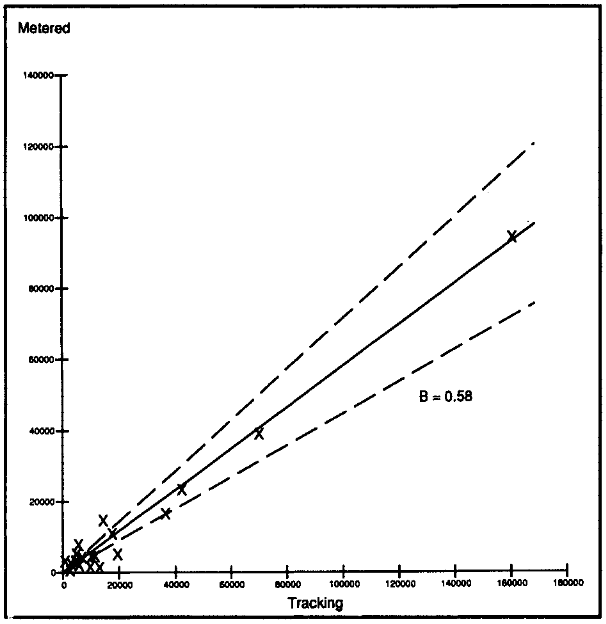

Figure 4. Actual kWh Savings

projects would still give ±20%, which may be adequate for many purposes.

ECA Using Double Sampling

The Company is very interested in ways of improving the engineering estimates to ensure an adequately small error ratio. The benefit of accomplishing this would be reflected in smaller required end-use-metered samples and better precision. However, there are limitations in the quality of the data that can be collected cost effectively for all projects in a program. For further improvements in the error ratio, NU is considering various ways of supplementing the tracking information with supporting information from other sources such as on-site audits, billing information, and load research information. These data can be analyzed using various methodologies such as PRISM billing analysis and statistically adjusted engineering (SAE) estimation.

The Company is employing a technique called *double sampling* to integrate the supporting information with end-use metering or other suitable data. The approach is to:

- Develop the on-site audits and other supporting information in a relatively large sample, called the first-phase sample or *supporting sample*.
- Prepare improved engineering estimates of impact for each project in the supporting sample using the on-site audits, site specific PRISM billing analysis, and so on.
- Use tracking information to extrapolate the results from the supporting sample.
- Use end-use metering and other appropriate monitoring as a bench mark to "true up" these results.

Table 2. Hypothetical Example of Double Sampling

Method	Savings (MWh)	Relative Precision
Average Savings:		
All Projects (N = 4764)	19.093[a]	
Supporting Sample (N = 200)	18.899[a] 12.851[b]	±5.8%
End-use Sample (N = 25)	11.089[b] 9.368[c]	±8.2%
Combining the Information:		
Estimated Total Savings: 4,764 • 19.093 • (12.851/18.899) • (9.368/11.089) = 52,251 MWh		$\sqrt{0.058^2 + 0.082^2} = \pm 10.0\%$
Confidence Interval: 47,026 to 57,477 MWh		

[a]From tracking information.
[b]From improved engineering estimates.
[c]From end-use metering.

In practice, on-site surveys would be carried out in a relatively large supporting sample and the end-use metering would be carried out in a much smaller sample nested within the supporting sample. Table 2 summarizes a hypothetical example in which on-site inspections are carried out in 200 projects and 25 of these projects are also end-use-metered. Table 2 illustrates how the resulting data would be analyzed.

In the supporting sample, the key statistic is the ratio between the average of the improved engineering estimates of savings to the average of the savings calculated from tracking information. An initial bias correction can be made by multiplying the total savings calculated from the tracking information by the ratio 12.851/18.899 = 0.6800. The end-use metering provides a second bias adjustment, determined by the ratio of the savings determined from the metering to the improved engineering estimates of the savings, 9.368/11.089 = 0.8448. The two ratios, 0.6800 and 0.8448, are used together to true up the engineering estimates available for all projects as shown in Table 2, giving an unbiased estimate of total savings of 52,251 MWh.

A simple procedure can be used to estimate the precision of the combined result. The first step is to relate the improved engineering estimates in the supporting sample to the engineering estimates from tracking information. Because the supporting sample is relatively large, the relative precision found in this step is quite good, ±5.8%. In the second step, metered savings are related to the improved engineering estimates. Since the improved engineering estimates give a small error ratio, the relative precision in this step is also quite good—about ±8.2%. The third step is to estimate the combined relative precision using the simplified calculation $\sqrt{.058^2 + .082^2}$. The combined relative precision is about ±10%.

The double sampling strategy is very effective in this example. Without the use of the on-site inspections, the 25 end-use-metered projects would have given ±20% precision. Or, under the conservative assumptions of the example, about 100 end-use-metered projects would have been required to achieve ±10% precision. With double sampling, the end-use metering can be limited to 25 projects augmented by on-site inspections in an additional 175 projects. These results are under quite conservative assumptions, namely that the overall error ratio is about 60% while the error ratio relating the metered savings to the improved engineering estimates is about 25%.

In most double sampling applications, the supporting sample is substantially larger than the end-use-metered sample, but much smaller than the total number of projects in the program. The Company is confident that substantially improved engineering estimates can be developed by concentrating on a moderately-sized supporting sample. By developing good predictors of the actual savings, the end-use-metered sample can be kept small without sacrificing statistical precision.

Concluding Remarks

Northeast Utilities is employing engineering estimates, billing analysis, end-use meterin,g and other suitable impact evaluation methodologies. NU is developing ECA as a technique for integrating the various sources of information to achieve consistent results efficiently. The early results are encouraging. The ESLR study indicates the potential of ECA at integrating end-use metering with tracking information and on-site inspections in a commercial lighting rebate program. The Company has additional studies under way to test the ECA approach in other C&LM programs.