MEASUREMENT OF ENERGY SAVINGS FROM DEMAND SIDE MANAGEMENT PROGRAMS IN U.S. ELECTRIC UTILITIES

Margaret F. Fels and Kenneth M. Keating*

PU/CEES Report No. 276

April 1993

*Bonneville Power Administration, Portland, OR 97232

Center for Energy and Environmental Studies
The Engineering Quadrangle
Princeton University
Princeton, NJ 08544

To be published in the Annual Review of Energy and the Environment, Volume 18

CONTENTS

I. INTRODUCTION

Demand Side Management is a big business in the 1990's. Measuring its impacts is both a big business and a challenge, politically as well as intellectually.

In popular parlance, Demand Side Management (DSM) means "conservation." Treating energy conservation as a resource, DSM allows a utility to control the balance of its resources and demands for energy by managing the consumer's needs for energy rather than by simply adding more supply. Although it can involve shifting loads, and even strategically increasing loads at certain times of the day, it has generally meant reducing the need for energy at particular times (demand savings), or reducing total consumption (energy savings), through efficiency improvements. The measurement of the energy or demand savings achieved by DSM programs is the conservation "meter".

This paper reviews the efforts in the U.S. to measure the effects of programs that try to conserve energy and/or manage demand. The review concentrates on the evaluation of efforts in electric utilities, because that is the focus of the vast majority of the literature and experience.

Based on an analytic approach that was developed in the 1930's for agricultural programs and expanded during the 1960's to cover social programs, program evaluation was adapted to energy programs in the late 1970's. The early efforts were directed at Carter-era Federal programs, such as the State Energy Conservation Program (SECP) and the Energy Extension Service. These tended to be ad hoc evaluations, which over time have become more organized and inclusive [see historical perspective by Hirst (1)].

The original purpose of program evaluation was to see if a program effect could be detected. Simply seeking to determine if the early programs produced statistically significant results was considered a challenge when the programs were emphasizing information and education. As conservation programs became more substantial, with incentives offered for physical improvements in efficiency, measurement of the size of the impact became the evaluation emphasis. Today, with very large utility investments in DSM and with DSM seen as part of a utility's resource portfolio, evaluations require more precise and reliable estimates of the exact magnitude of the savings.

The field of DSM evaluation has evolved over the last 15 years. By the late 1980's some utilities were beginning to consider evaluation as an integral part of good program management. Their interest in evaluation grew as they recognized that DSM programs were important customer service vehicles; they wanted to know whether they were providing good service. With the recognition of DSM as a viable energy resource alternative, interest in measuring energy savings increased. However, only the decision of many state regulatory agencies to grant shareholder incentives based on measured program results has given evaluation its current high profile (2). [As of March 1992, 21 states were giving this type of incentive (3).] In the current regulatory environment, DSM is regarded as a resource, and the results of the evaluations have real monetary impacts on utilities.

In some states, evaluators are still trying to establish the need for evaluations; in other states, decisionmakers are making demands of evaluators for levels of precision and breadth that are not consistently achievable. In particular, evaluators have been "hoisted upon their own t-test." After years of trying to sell the usefulness of evaluation and its scientific nature, evaluators may have oversold the field. An expectation that has been growing is for measurement techniques to begin isolating the effects of individual measures over a representative sample. As with the need to measure coincident and non-coincident demand impacts, this will require greater emphasis on specialized metering. One of the lessons made clear in this review is that, just as the field has evolved over the last decade, it will continue to evolve, along with the goals, methods, research designs, and metering technology. The multitude of references used in this review, selected to represent the range of evaluation approaches and in no way exhaustive, attests to the rapidly changing and boldly experimental character of the field.

One thing that is unlikely to change is the desire to measure the impacts that are attributable to the program or utility intervention. This issue of attribution, or net savings, is central to the methods and results reported in this paper. The objective of an evaluation is not to measure changes in consumption, but, rather, the changes that are attributable to the program. Much of the research has been devoted to trying to isolate program effects from those of weather, electricity price, economic cycles, consumer behavior unrelated to the program, and background variances in consumption. Any methodology that does not address the issue of attribution adequately will be criticized, but no one methodology is the solution to the problem. Sometimes multiple methods must be used, and sometimes highly accurate attribution with the desired detail is simply not feasible.

In order for the DSM resources to displace supply alternatives successfully, the impacts of the DSM programs must be long-lasting and reliable. Nevertheless, notwithstanding the considerable research to date on measuring energy savings, the measurement of the persistence of such savings is relatively limited (4). The study of persistence is now an important initiative in the evaluation field, as summarized in two recent reviews (5, 6).

In spite of the need for long-lasting savings, contractual agreements in which payments are tied to estimated energy savings put pressure on evaluators for quick answers. Inevitably, accuracy will be sacrificed for timeliness. This need for "verification" savings places evaluation in a very different context from one in which the objective is to learn as much as possible from the program at hand, to assess its cost-effectiveness and to help program designers decide which DSM approaches to emphasize in future programs. [Requirements for evaluation vs. verification are compared in Kushler et al (7)]. Ideally, an evaluation plan would involve verification followed by a full-scale evaluation, with the verification methods designed as short-cut versions of the full-scale methods. In actual DSM evaluations to date, the distinction between verification and evaluation is generally blurred. In reviewing techniques for measuring energy savings, this paper will emphasize methods designed for full-scale program evaluation, although the same methods are generally useful for both verification and evaluation.

Often the greatest value of DSM is that it helps a utility avoid or delay the construction of new capacity, and sometimes upgrades to its distribution system. In order to do this, the DSM programs must reduce the demand for energy at specific periods of high demand, as in a winter evening or a summer afternoon. Measures such as thermal or cool storage help shift loads from high-use to low-use periods. Most conservation programs result in some valuable demand reductions due to efficiency improvements that result in a lower average demand for energy at the key times for the utility. The techniques for measuring these effects can be expensive and have been far less widely applied than their counterparts for measuring energy savings. They will be addressed briefly.

The paper will begin with an overview of measurement methods that will be discussed in the remainder of the paper (Section II). The measurement experience in residential, commercial/institutional, and industrial programs will then be summarized (Section III). These discussions will address both retrofit and new construction programs. Drawing from specific evaluations, generalizable lessons will be emphasized. The paper will end with a discussion of where the field is likely to find its future challenges (Section IV).

II. DESCRIPTION OF METHODS

II.A. The Crystal Ball Problem

When an evaluator "measures" energy savings from a DSM program, he or she is trying to estimate the difference between the amount that would have been consumed in the absence of the program and the amount actually consumed, i.e.,

$$\text{Savings} = (\text{kWh})_{\text{"would have"}} - (\text{kWh})_{\text{actual}} \quad . \qquad (1)$$

A complete evaluation toolkit would include: metered electricity consumption data (whole-building and/or end-use specific) from before and after installation of the DSM measures, an engineering understanding of what end uses are likely to be affected, data on building characteristics, and a crystal ball for what would have been consumed. Since the latter is not available at any price, much of the challenge for evaluators has been to develop techniques, using metered data, engineering estimation, and information about non-participants, to replicate the crystal-ball function.

A first approximation to the amount the participant "would have" consumed is the actual amount consumed before the DSM program. Adjustment for differences in weather between the year after and the year before the program is essential so that program effects are neither masked nor exaggerated by changes in weather. The weather-adjusted savings may be written as:

$$\text{Savings} = (\text{kWh})_{\text{adj,pre}} - (\text{kWh})_{\text{adj,post}} \quad . \qquad (2)$$

In addition to weather, confounding factors such as energy price and attitude about conservation could cause a change in consumption independent of the DSM program. A comparison group of non-participants can be used to control for these effects. The net savings for the group of DSM participants becomes:

Net Savings = Savings (DSM participants group) - Savings (comparison group). (3)

Further adjustment for differences between the participant and comparison groups is possible by including variables such as house and household size in the analysis.

A nagging issue with attribution in DSM evaluations is "free ridership." A free rider is a program participant who would have taken the program action even if the program did not exist. Because the savings would have, by definition, occurred without the program, the inclusion of the savings due to free riders inflates the savings. To the extent that a comparison group of non-participants adequately reflects what the participants would have done in the absence of a program, a design that uses comparison groups can net out the effects of free riders along with other confounding factors. However, the validity of this design becomes harder to support in cases where the non-participants are taking actions as a spillover effect of the program, due to example or program advertising. Thus the comparison group technique may not identify the exact free-rider contribution, as required by some regulators. This problem is receiving considerable attention in evaluation (8), as a major confounding issue.

In theory, the process of estimating net savings can apply to savings at the whole-building level or for isolated end uses, for programs in any sector, and for new construction as well as retrofit measures. In practice, this process has mainly been applied to the measurement of whole-house savings in residential programs. In new construction, absence of a "pre-period" means that the evaluator needs to find a different measure of what would have been consumed. In programs of large and disparate commercial or industrial buildings, well matched comparison groups are not practicable or meaningful, yet the need to adjust for other factors may be even greater in large buildings with high process loads and varying levels of business than it is in residential buildings. When a DSM program is focused on a specific end use, such as lighting, and especially when the effect is thought to be not clearly discernible at the whole-building level, the evaluation is likely to want to focus on estimation of savings for the end use in question; yet sub-metered end-use consumption data are costly, pre-retrofit sub-metered data are usually unavailable, and the possibility that lower consumption in the metered end use (e.g., lighting) could increase consumption in an unmetered end use (space heating) may be overlooked (9). These are among the many dilemmas facing the evaluator as he or she reaches into the kit of available tools for tackling the crystal-ball problem.

Methods for estimation and measurement of energy savings are strongly characterized by their data requirements and the extent to which they are based on engineering estimates or metered consumption. The spectrum of possible methods is broad indeed, spanning the gamut from overly simplistic look-up tables of expected percent savings, with the "data" consisting of conservation actions taken in each building, to impractical sub-metering of each appliance or end use expected to be affected by the DSM program, for non-participants as well as for participants, and for a sufficiently long period before and after the retrofit to allow adjustment for seasonal effects. In between these extremes is a variety of methods that evaluators are finding both useful and practical, though limited in different ways. In this review, these are described in four categories: 1) engineering estimation techniques; 2) billing analysis tools based on whole-building metered data; 3) multiple regression models based on billing and survey data; and 4) methods based on end-use sub-metering. In addition, combined approaches are discussed, as a possible route toward compensating for the limitations of individual methods.

II.B. Engineering Estimation

Since the earliest conservation programs, engineering modeling has played a key role in their planning and implementation. In almost all programs, such models are used to predict the energy (and demand) savings from specific measures, at the individual-building and program levels, and thereby to identify the most effective strategies for saving energy. Given their availability, it is not surprising that engineering models have been a popular method for estimating program impacts. However, as borne out by experience, a prediction is often a flawed representation of what actually happened.

The most highly developed engineering models are building simulation models, such as DOE-2, which involve a detailed modeling of a building's heat flow and its response to outside weather. (See 10 for a comprehensive and up-to-date summary of simulation models; see also 11, Chapter 4.) The DOE-2 model was fully developed in 1980 (12), and more recent models such as ASEAM (13) and SUNDAY (14) have built on that experience. Because they require detailed data on building and occupancy characteristics, simulation models are costly to run, but they are also amenable to sensitivity analyses of the response of a building's energy consumption to specified changes in its characteristics (added insulation, reduced wattage in light fixtures, etc.).

Much simpler and less costly engineering estimation procedures, appropriate to straightforward spreadsheet analysis, are frequently applied in DSM program evaluation. An example of a simple engineering algorithm for the annual energy savings from a high-efficiency lighting program is as follows:

$$\text{Savings in lighting} = \Sigma_i \left(kW_{pre}^{\,i} - kW_{post}^{\,i}\right) H_i \qquad (4)$$

where H_i is the number of hours of expected operation per year for each type (i) of fixture retrofitted. Well suited to predicting energy savings, engineering models such as this are often adopted as short-cut methods for estimating after-program savings. For instance, if knowledge of the energy savings of a specific end use is desired, as is increasingly the case with DSM programs in large buildings, and if metered consumption is required only "where practicable" [in the words of a New Jersey rule-making (15)], engineering estimation is likely to be used. In addition, it is often considered the most practical method when estimates of energy savings are desired that are independent of the influence of confounding factors such as changes in occupancy or uses of a building.

Unlike the other approaches discussed in this review, engineering methods are not limited by the availability of metered consumption data. While this feature makes them appealing, in that in theory any question can be answered by the "right" engineering model, it is also the Achilles heel of engineering estimation because many factors, such as behavioral or occupancy changes, or low-quality installation of the measures, can erode the accuracy of predictions of energy savings. From evaluations to date (as discussed in Section III), there is considerable evidence that engineering-based estimates of energy savings are often inaccurate, and in general overstate the actual savings achieved. Because of the broad flexibility of engineering models, as well as their simplicity and low cost, engineering models are often incorporated into a broader evaluation approach, wherein they are calibrated by some post-program measured data, or they are used in combination with statistical methods based on metered data (see Section II.F).

II.C. Billing Analysis

Practically all buildings that participate in DSM programs have monthly meter readings representing whole-building electricity consumption. Similar data are available for buildings that might serve as a comparison group. The fact that month-to-month variability in a building's consumption is highly explainable by fluctuations in outside temperatures makes simple modeling of billing data feasible. Knowledge of the temperature sensitivity of space heating (and cooling) in the utility world is certainly not new, and can be traced back in the literature to the early 1900's (16, 17). Simple temperature models are often used for weather adjustment of consumption to correct for the effect of more severe or more mild weather in post-program vs. pre-program consumption. The simplest approximation is to regress the billing data against degree-days, and rescale to degree-days in a typical year.

The most widely used billing analysis tool is PRISM (PRInceton Scorekeeping Method) (18), which differs from other weather adjustment techniques in its inclusion of a variable reference temperature as the degree-day base (19) and in its extensive statistics that yield accurate error bars for the model estimates (20). Applied to consumption data from a building's energy bills and daily temperature data from a nearby weather station, PRISM determines a weather-adjusted index of consumption called Normalized Annual Consumption (NAC). The building's energy savings are then derived as the difference between the NAC in the pre- and post-weatherization periods, i.e.,

$$\text{Savings} = NAC_{pre} - NAC_{post} \quad . \tag{5}$$

Through the simplicity of the model, the savings estimates remain closely related to the original metered data and thus to the true consumption picture, ensuring transparency of the method.

Based on readily available data, billing analysis is appealing because of its widespread applicability: it can be applied to nearly all customers participating in a DSM program, as well as to a sample of buildings selected as a comparison group. The use of billing data has been tightly linked to quasi-experimental designs in program evaluation (21). This fairly standard research design, sometimes referred to as a "before-and-after comparison group design", uses a well matched comparison group of eligible non-participants to try to account for non-program effects. A snapshot of sample results from this type of design is provided in Figure 1. In this example, an evaluation of a Wisconsin low-income program by Goldberg (22), distributions of weather-adjusted whole-building energy savings [using Eq. (5)] in the treatment vs. comparison groups are compared to determine the net savings [Eq. (3)].

Through the conservation programs of the 1980's, the feasibility of using billing data for the estimation of whole-building energy savings has been demonstrated by the use of PRISM in numerous large-scale evaluations of conservation programs (examples are reported in 22, 23, 24, 25, 26). One advantage of such widespread use of a single method is the possibility of meaningful comparisons across programs, to shed light on what works and what doesn't work in conservation approaches. A good example is the Buildings Energy-Use Compilation and Analysis (BECA) data base, an ongoing project that compiles measured data on the effectiveness of single-family retrofits, and reports PRISM estimates wherever possible (27).

Most billing-analysis evaluations have been of residential programs. The applicability of PRISM to DSM programs in commercial buildings is currently being explored, in research (9, 28, 29), and in actual measurement of savings (30, 31, 32). The results of these evaluations show the potential usefulness of billing analysis in nonresidential buildings, and in general in larger buildings that are increasingly the focus of DSM programs, although many questions remain (33).

II.D. Multivariate Regression Analysis

When survey data are available on customer characteristics for participants and a comparison group of non-participants, cross-sectional multivariate regression (MVR) modeling using these data in combination with monthly billing data is a popular tool for inferring program-attributable energy savings from the data. Model specification, which is strongly influenced by data availability, may be written as:

$$F_i = \alpha_1 + \alpha_2 X_{i2} + \alpha_3 X_{i3} + \ \dots\dots \ \alpha_n X_{in} + \varepsilon_i \ . \qquad (6)$$

where F_i denotes the dependent (fuel or electricity consumption) variable for the i^{th} observation (participant or non-participant), X_{ij} the j^{th} independent or explanatory variable for the i^{th} observation, α_j the corresponding coefficient to be determined by the model, and ε_i the error term (see 11, Sections 5.2-5.5, for formulation of MVR techniques).

Generally, independent variables include some representation of heating and/or cooling degree-days, program participation (a dummy variable, which is 1 if household or building participated and 0 otherwise), as well as information on space conditioning and appliance ownership (e.g., whether electricity for the i^{th} participant is expected to be used for space heating, air conditioning, etc.). In addition, interaction variables made up of products of other variables, such as program cost and heating degree-days, are often added (for examples, see 34, 35, 36, 37). Inclusion of interaction (and other) variables may be justified on the basis of simple statistical tests for significance, but ultimately "determining whether interaction variables should be used depends primarily on the researcher's judgement" (11, p. 5-20).

One choice of dependent variable is consumption for the year after implementation of the conservation program, with the coefficient of the participation variable interpreted as the savings attributable to the program. This approach does not include pre-program consumption data. Another approach is to use the pre/post difference in consumption as the dependent variable, again with the coefficient of the participation variable interpreted as the energy savings. Such choice of variables can both influence interpretation of the model results and also dramatically affect the model's statistics on which the credibility of the results is often judged (see discussion of two-stage PRISM in Section II.F.4).

The most systematic modeling approach in this category is Conditional Demand Analysis (CDA), which was developed by Parti and Parti to estimate appliance-specific energy consumption from electricity billing and appliance ownership data (see 38, 39 for description of the methodology, and 37, 40 for sample applications to DSM program evaluation). CDA models differ from other, more ad hoc, MVR approaches in their incorporation of engineering principles in addition to demographic variables in the model structure. When a particular end use is of interest, recent CDA models have imbedded engineering models of that end use into the regression equation, to reflect, for example, the expected dependence of water-heating energy use on the difference between the temperature of the hot water in the heater and of the ground water. This makes the interpretation of the model coefficients more meaningful. Nevertheless, CDA models that fail to include comparison groups are not convincing in their attribution of savings to the program.

Those features of multivariate regression that make it so appealing for DSM evaluation are also its potential pitfalls. Flexibility of model specification allows the evaluator to include end uses and other variables whose coefficients are of interest, and also to base the choice of variables on data availability. The all-in-one approach, wherein the same model is used to estimate savings while controlling for as many non-program factors as possible that may influence energy consumption and energy savings, is also appealing. However, the flexibility of variables can give evaluators the misleading impression that, by choosing a certain set of variables, accurate attribution of savings to the DSM program has been achieved. In the real world, the results of even the most carefully conceived regression models are inevitably sensitive to the variables included. Furthermore, the all-in-one feature carries with it the danger of a "black-box" approach whose results and implications are far from transparent, and for which guidelines to ensure unbiased evaluation are difficult to establish. In our less-than-perfect modeling world, complex relationships between energy consumption and human behavior and other factors are non-quantifiable (as well as little understood), making the "right" choice of variables far from obvious.

II.E. Methods Based on End-use Sub-metering

Billing analysis and multivariate techniques, discussed above, depend on metered data as a reliable and objective measure of consumption, usually at the level of an entire building, and typically with monthly frequency. Sometimes more disaggregation and/or more frequent measurements are needed, which calls for special end-use metering, or sub-metering.

Detailed sub-metering is an alternative when the energy savings expected from a measure represent a very small fraction of the total energy use of a building or set of buildings, and when the sample sizes are too small to expect statistical techniques to isolate the savings. An example would be the effect of lighting measures in an industrial plant, or when only a small portion of a major office building is affected. The savings can be large in absolute terms, but small compared to the natural variation in consumption. Specialized industrial measures and site-specific program approaches usually require specially designed metering strategies. In addition, because many utilities place priority on the savings that occur at a particular time of the day or week, frequent monitoring of loads becomes important. Metering technology permits measurements to be as frequent as many times per second, but fifteen-minute or hourly data recording is fairly standard.

End-use metering for DSM evaluation builds on experience from utility load research. Examples include Bonneville Power Administration's End-Use Load and Consumer Assessment Program (ELCAP) project (41). [See excellent review of end-use load shape research by Eto et al (42), and summary of relevant technology by Misuriello (43).] Sub-metering doesn't only refer to measurement of Watts or kiloWatts. Measurements can include temperature, air or fluid flow rates, cycling of specific equipment, or other parameters necessary to isolate or extrapolate the energy savings. In the Hood River Conservation Project, for instance, indoor temperature and the radiant output from wood stoves were measured every 15 minutes for several years (44). The needs of researchers have driven the technology in this area, and even efforts to catalogue the equipment available are outdated within months (45).

The high cost of sub-metering has been a hindrance to its use. Early efforts cost upwards of $20,000 to monitor a single commercial building exhaustively for a year. When researchers can restrict the measurements to specific circuits or pieces of equipment, costs have been greatly reduced: a lighting circuit can be monitored on a fifteen-minute basis for less than $400, providing detailed load shape information. In addition (as discussed in Section II.F.2), cost reductions are possible by reducing required sample sizes through more efficient sampling techniques (46).

Cost is not the only barrier to the use of sub-metering. Attribution, the capture of interactions among DSM measures, and extrapolation from short periods of metering to annual savings are other difficult issues in sub-meter-based evaluation methods. The use of comparison groups to identify net savings, even if available sample sizes are large enough and appropriate comparisons can be found, can double the cost of sub-metering. Quasi-experimental designs are, therefore, seldom feasible. Other techniques, such as turning the DSM measure on and off (47) or controlling for industrial input and output (48), have been successful.

An alternative approach requiring similar data acquisition technology is to infer changes in end-use load shapes from whole-building load data (49, 50, 51). For large buildings, hourly load data before and after program installation may have been collected by the utility for billing purposes. Provided the effect of the DSM intervention is discernible in the whole-building load shape, this approach may be a particularly useful way to account for interactions among DSM measures and between the measures and other energy systems, such as the effect of reduced lighting loads on heating and cooling needs, which are usually ignored in sub-metering (52; see also 9).

To keep costs down and to keep historical trends from confounding the measure of changes in consumption, sub-metering is sometimes made on a short-term basis (47, 52, 53). However, this restricts the ability to estimate annual savings. Two weeks of sub-metering before and after the installation of measures is basically an opportunity sample in time, with a schedule determined by the timing of measures rather than by evaluation design. Detailed data on spring usage and operating hours may not reflect peak summer consumption in an office or the operating schedule in a retail store during the holiday season, and interactions between end uses can vary by season also.

Despite the identifiable problems, sub-metering is expected to play an increasing role in the measurement of DSM impacts (7, 54). The role is expected to include sub-metering in combination with statistical approaches to measurement (55), as discussed below.

II.F. Combined Approaches

Evaluation approaches often combine the features of more than one method, generally with the objective of correcting for the deficiencies of one method with another. Several combined approaches are described here.

II.F.1 Statistically Adjusted Engineering Models

The main pitfall of engineering estimation is the lack of connection with actual (metered) consumption. Alternative approaches based on metered data reflect what actually happened, but often not with the desired end-use (or other) detail. Developed as a response to both problems, Statistically Adjusted Engineering (SAE) models use metered data in combination with engineering models to calibrate the prior engineering estimates of savings or to provide end-use disaggregation of whole-building metered consumption. When the engineering estimates are based on simulation models, an approach called simulation tuning is used.

A specific approach, the Hybrid Statistical/Engineering Method (HSEM) for estimating load impacts as well as energy savings, uses engineering estimates of end-use energy savings (or load reductions) for each end use, and metered whole-building consumption (or hourly metered load data, for the load-reduction model) for a comparable period (56). In an MVR format, the independent variables might include: engineering estimates of reduced energy use for space heating, water heating, and other end uses; degree-days; and terms for economic and behavioral variables. The coefficients of the engineering estimates provide an indication of how much each corresponding estimate was over- or underestimated. (For an application of this type of SAE model, see 57; for a general discussion, see 11, Section 7.4).

II.F.2. Engineering Calibration Approach

In programs with sub-metering, end-use metered data are generally available for only a small fraction of the participating buildings, and may span a relatively short time interval. In a program in which engineering estimates of end-use savings have been calculated for all program participants, for example, during the design stage, metered end-use data for a small sample can provide "a benchmark to true up the engineering estimates" [in the words of Amalfi and Wright (58), p. 194]. In general, combining auxiliary data and other estimates with end-use metered data is a way of extending the usefulness of sub-metered data.

An example is the Engineering Calibration Approach (ECA), which is designed to extrapolate specific end-use metered estimates of savings, for selected measures within a sample of sub-metered buildings, to the entire program. To do this, end-use metered data from the sample of participants are integrated with engineering estimates for all of (or a large number of) the participants and other available data such as on-site surveys, SAE estimates, and billing data, using sampling techniques such as Townsley and Wright's Model Based Statistical Sampling (MBSS) (46, 59). Cost savings from applying such techniques are evident (60). However, they generally do not include non-participants, and thus their usefulness for net savings is limited.

II.F.3 Short-term Methods

Billing analysis approaches discussed thus far require a full year of post-retrofit data and thus more than a year between the program and evaluation results. Evaluators are sometimes pressured to provide earlier estimates of savings. In response to this, short-term estimation techniques, which draw on billing analysis and end-use metering, have been developed. One short-term technique uses several months of post-treatment billing data, and a year of pre-treatment billing data modeled by PRISM to predict what consumption in the same post-treatment period would have been without the program (61, 62). Another approach uses sub-metered runtime data (e.g., on furnaces) for a short period before and after the treatment, generally in mid-winter so the monitoring is completed in the same season (63).

Although the desire for prompt evaluation results is understandable, possible bias from use of data spanning only a fraction of a year has been identified (64). When short-term methods are required to provide early indicators of energy savings, a follow-up billing analysis that includes a full year of post-retrofit data can provide a check on whether the short-term estimates can be extrapolated to annual savings.

II.F.4. Two-stage PRISM

A billing analysis approach such as PRISM that uses a comparison group to adjust for what the program participants "would have done" in the absence of the program assumes that the comparison and participant groups are well matched. When survey data on household or building and occupancy characteristics are available, an MVR analysis may be added as a second stage to the billing analysis, in order to adjust the savings estimates for differences between non-participants and participants (such as self-selection bias) as well as for other factors. The main difference between two-stage

PRISM and an all-in-one MVR approach is the incorporation of the best possible (non-linear) physical model relating consumption to weather, as stage one in the former, and the inclusion of the interaction between weather and other variables, in the latter.

Two-stage PRISM has been used in a number of large-scale evaluations of residential programs, and was first applied by Oak Ridge National Laboratory to evaluations of Bonneville Power Administration (BPA) programs (65, 66). It has also been found to be useful in evaluations of energy-efficient new home construction (67).

In the two-stage approach, the distributions of weather-adjusted savings across houses (or buildings) in the participant and comparison groups, with statistics concerning reliability and significance of their means and medians, comprise "stage one" of the analysis (Figure 1). "Stage two" consists of combining the individual-house savings estimates with other variables in a cross-sectional model. Independent variables for a residential program might include household income, number of household members, floor area, electricity price, whether a secondary fuel was used, as well as program participation.

One issue in the construction of the stage-two model, mentioned earlier when MVR was introduced, is whether weather-adjusted savings or consumption should be used as the dependent variable. Since the variability of savings across households is of primary interest, the savings estimates may be the better choice, possibly with pre-program consumption included as an explanatory variable. Such decisions can have a dramatic effect on the R^2 of the cross-sectional model, whose closeness to 1.0 reflects how well the specified model fits the data. In one evaluation (66), several equally meaningful stage-two models were compared, with a large range in R^2 values (from 0.5 to 0.8). Realistically, this sensitivity of R^2 to the variables selected can pose difficulties for the evaluator, whose results may be judged by regulators and program managers as

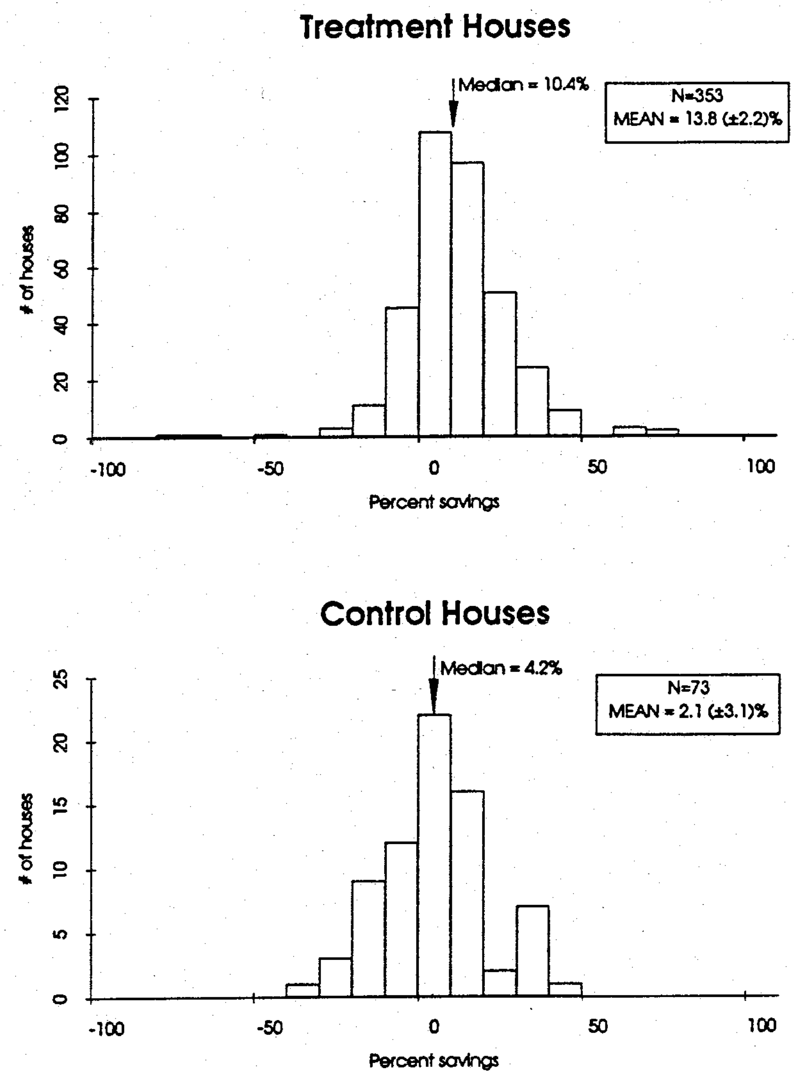

Figure 1. Snapshot summary of a sample program evaluation (from 22, p. 41). Weather-adjusted savings are estimated by billing analysis using a quasi-experimental design. Distributions of percent savings of the treatment vs. comparison group of houses are shown. Figure summarizes stage one of two-stage PRISM.

more believable if the R^2 reported for the evaluation model is high, regardless of its interpretation.

Many evaluators find two-stage PRISM intuitively appealing because, in contrast to a one-stage MVR approach, it preserves the transparency available in a PRISM analysis, in stage one, at the same time including the effects of additional variables typically used in MVR or CDA models, in stage two. Even though the second stage often causes only a minor adjustment in the savings estimates (65), its inclusion enhances the understanding of the net savings and the validity of the comparison group used.

II.F.5. Discrete Choice Models

Two issues in impact evaluation that may strongly influence an evaluator's choice of modeling approach are self-selection bias, from differences between the participant and comparison groups, and free ridership, reflecting participants who on their own would have taken the conservation action promoted by the program. A perfect, well isolated comparison group, designed to estimate how much the participants would have consumed in the absence of the program, in theory would account for both of these effects. Although free riders contribute to realized program savings, understanding their contribution to savings is an important attribution issue for utilities. Free ridership is one type of self-selection bias that can show up in a net savings analysis, provided the self-selection bias is not strong.

A class of models designed to correct for self-selection bias is Discrete Choice (Participation) models. Applied to DSM program evaluation, they build on experience from other fields such as consumer purchases and job preferences (see 68 for background on discrete-choice theory) and use a variety of mathematical modeling tools building on probability as well as regression theory (11, Sections 2.6.4 and 6.0).

Although Discrete-Choice models are used to estimate adjustment factors for energy savings rather than to estimate energy savings directly, they are included in this review, both for their usefulness in conjunction with the savings-estimation models and for their potential enhancement of the interpretation of the savings estimates.

II.G. Evaluating Peak Demand Savings

Knowing exactly when savings in demand occur, and at the same time attributing them to the program intervention, is a daunting goal. Time-of-use metering on the affected end uses, and those with which they interact, both before and after program intervention, is very expensive, and prohibitively so if carried out for a long enough period on a fully representative sample with an appropriate comparison group. Yet demand impacts are very important to quantify. As with program evaluation in general, researchers have taken approaches that recognize that full knowledge isn't likely and that all evaluation is an estimation process. This has resulted in a menu of options that partially parallel those for energy savings.

Engineering simulation models can be used. The more sophisticated ones allow the analyst to look at simulated loads by time of day, controlling for outside temperature, before and after DSM measures are installed (10). Adjustments can be made for confounding factors such as free ridership or economic activity, to approximate net demand effects. The results may or may not reflect the reality in the field. Whether based on complex simulations or simple engineering algorithms, this may be the only possible approach short of sub-metering for DSM efforts in large commercial and industrial buildings.

An easily applied approach takes the energy savings estimated with any of the methods detailed earlier, or net savings when available, and applies load research models to assign load shapes to the savings (53, 69). Since these models use previously established correlations between energy sales and loads at specific time periods, an intrinsic problem is the assumption that the DSM measures will have the same effects on loads as temperature, economic activity, and energy prices that drive the historical models.

When the value of demand savings is determined by a few major annual and daily time periods, some utilities use "hours of use" surveys to identify the time of day a measure is in use (47). If the surveys indicate, for instance, that residential lights are not on between noon and 8 PM in August, the evaluators estimate that there was little or no peak demand savings from compact fluorescent bulbs. If 90 percent of commercial lights are on during the same peak period, then 90 percent of the estimated or measured demand reduction from commercial lighting is ascribed to peak-load reduction.

Sub-metering from small samples of participant buildings, for short periods before and after participation, may be applied in buildings with reliably known schedules. The resulting observed percentage of lighting fixtures that are on at the time of day at which peak demand usually occurs is then used to estimate the demand savings on peak (31). Further adjustments are necessary to approximate net program impacts.

One potential approach that has not been tried, to our knowledge, is to identify program participants that are also in a utility's load research sample and compare the hourly data from the two data sets. With good before- and after-program data and a properly selected comparison group from among the other load research buildings, net whole-building demand effects could be estimated. A practical disadvantage is that end-use level data will not be directly available. Despite the analytical problem that changes in end uses not impacted by program measures may confound the analysis, an advantage to this approach is that the interactions among measures and other end uses would be captured.

Whether or not the same method forms the basis for estimating demand and energy savings varies widely among evaluations. In theory, energy consumption (kWh) can be derived from demand (kW) data integrated with hours-of-use (h) information, for specific end uses or the building as a whole. In practice, however, given the limitations of each evaluation approach and the challenge of answering complex questions with available data, the method used for estimating demand savings may not be the most appropriate for energy savings, or vice versa. For example, some form of hourly data is needed for the measurement of demand savings, whereas program-wide energy savings may be better estimated from billing data for a large sample of participants and non-participants than from a limited sample of sub-metered data.

III. LESSONS FROM MEASUREMENT EXPERIENCE

Experience in the measurement of energy savings varies widely by sector, both because utility programs targeted at different types of customers have different histories, and because the measurement issues faced by each sector influence the level of complexity and input data defining the evaluation approach. In the last decade, the number of evaluations that offer lessons of experience has mushroomed. The increased participation in the biennial International Energy Program Evaluation Conference, from 125 in 1985 to over 400 participants in 1991, is an indicator of the extent of this growth. The evaluations presented here are examples selected to illustrate not only a variety of applications but also problems and real-world challenges faced by evaluators.

III.A. Residential Buildings

III.A.1. Residential Retrofit Programs

Although accurate measurement of energy savings attributable to a residential program is far from easy, there is a longer history of evaluations in the residential sector than in any other, and there is more uniformity among evaluation approaches than in the other sectors. Because there may be thousands of households in a residential program, the evaluation often consists of billing analysis or MVR applied to a subsample of the participants and a comparison group. Common to most residential evaluations is the use of monthly metered data from energy bills.

A prototype evaluation of this type (single-stage PRISM) was of Princeton University's Modular Retrofit Experiment, a research project performed in the late 1970s (70). Some of the earliest evaluations of utility conservation programs were performed by Oak Ridge National Laboratory to evaluate programs in Connecticut and Minnesota (71, 72). Subsequently, extensive evaluations of BPA weatherization programs contributed significantly to the development of two-stage PRISM (66). In the evaluation of the $21 million Hood River Conservation Project, furnace sub-metered data and auxiliary data on indoor temperatures and wood use helped the interpretation of the results (23, 44).

Some large-scale evaluations have used multiple approaches, depending on data availability and type of program being evaluated. In the ongoing PRISM-based evaluation of the U.S. DOE Weatherization Assistance Program (WAP), single-family homes heated by gas or electricity, the largest component of the evaluation, are being evaluated with a carefully constructed nationwide sample of participant and comparison houses (24, 73). The fuel oil component, for WAP participants with oil heating, is using a short-term metering approach (with mid-winter weatherization) (74), and the multifamily component is using an approach similar to the single-family study, but with alternative comparison-group designs and with supplemental site-visit data for selected buildings (75).

One of the problems faced in the WAP evaluation, and other evaluations especially of low-income programs, is sample attrition from bad or insufficient billing data. An hypothesis that such attrition leads to biased estimates has been studied (76), but with small samples. Ironically, the fact that savings cannot be reliably estimated for houses with too little data is simultaneously the reason for the problem and the barrier to analyzing it. More research is needed to understand the extent and direction of this bias, if there is any, and to explore the reliability of practical remedies, such as the inclusion of all observations in the sample averages and more sophisticated approaches for pooling the observations (77).

A repeated lesson from these and other meter-based evaluations is that actual savings (estimated from metered data) often fall short of expected savings (predicted from engineering models) (78) . One of the earliest "disappointments" was the evaluation of a Connecticut audit program (71), in which savings were much lower than expected. In the Hood River Conservation Project, the "bottom line" was that the actual savings were only 43% of those predicted by careful engineering estimates (23). In a comparison of two-stage PRISM with CDA in a Maine program, the two methods gave savings estimates that were closer to each other than to the engineering estimate, and, in both cases, considerably below the engineering estimate (by about 30%) (79).

As illustrated in Table 1, this large discrepancy between predicted and actual savings, and more specifically the large two-to-one ratio of predicted-to-actual savings, is not atypical of residential conservation programs (80). The models may have failed to capture factors such as displacement of secondary fuel use by electricity, poor installation of retrofits, unpredictable human behavior effects, and inaccurate assumptions by the auditors. In the first evaluation of the Wisconsin low-income

Table 1. Summary comparison of engineering estimates vs. actual meter-based estimates of savings for residential retrofit programs

Program	Program Description	Actual/Engineering (%)
CMP Energy Management Assistance	Low-income grants	40
CMP Pay As You Save	Utility grant and loan	47
CMP Energy Management Rebates	Rebates	15
CMP Packaged Weatherization	Standard weatherization package	36
CMP Weathershield	Low-income grants	22
GPU RECAP	Performance contracting	22-44
NU Performance Contracting	Performance contracting	22
BPA Residential Weatherization	Comprehensive weatherization	40-58
Hood River Conservation Project	Comprehensive weatherization	43
SCL HELP (multiplex)	Comprehensive weatherization	117
NEES Partners - Residential	Comprehensive weatherization	107

Source: Nadel and Keating (Ref. 80, p.25) program, in which billing analysis was enhanced by simple correlations of the results with other variables such as household characteristics, the lower-than-expected savings were attributed in part to quality of weatherization work (22). In a recent evaluation of a residential high-efficiency lighting program, for which a very simple engineering model of kWh savings could be expected to be accurate if the hours of operations were known, actual savings from a billing analysis were found to be only half of the predicted savings (81), because of inflated hours of use in the engineering estimates and the removal of lights by dissatisfied participants. Such evaluations, albeit bearers of bad news for utilities counting on achievement of predicted savings, provide convincing evidence of the need for actual, metered data in the measurement of energy savings.

III.A.2. Residential New Construction

Though relatively recent, evaluations of new residential construction programs, which focused on heating and cooling only, have been completed and accepted in Wisconsin (82), Maine (83), California (84), and the Pacific Northwest (67, 85, 86). They all involved the use of billing records for the new program homes and non-participating homes built to existing codes and practices, with varying statistical sophistication. Two of these evaluations (84, 86) include energy use simulation models as part of the overall evaluation, with special efforts to visit and record construction characteristics of a sample of both types of homes. The other two evaluations use a CDA approach to model whole-house consumption. In the MVR model constructed for the Maine evaluation, 20 variables were included. In addition, results of a discrete-choice model were applied to correct for self-selection bias. The resulting savings estimates range from 15 to 30 percent, in comparison with builders' estimates of about 25%.

The evaluation of BPA's Model Conservation Standards (MCS), based on two-stage PRISM, is another example of lower-than-expected savings: the shortfall in energy savings was between 25% and 50%, depending on the climate zone (67). The stage-two MVR model explained 39% of the variation in weather-adjusted consumption, with a high level of significance. Although evaluations of new construction programs have the disadvantage of no pre-program data, they indicate that meaningful "savings" are obtainable by careful comparison with non-program houses built at the same time.

III.B. Commercial and Institutional Buildings

The predominant trend in the measurement of net program effects for commercial retrofit programs has been toward increasing sophistication of techniques. Pigg has advocated this on the grounds of the perceived inadequacy of more traditional billing analysis and MVR in dealing with the complexities of large buildings (33). For the relatively homogeneous populations of small commercial buildings, a quasi-experimental design with some stratification has been found to be acceptable. However, for larger commercial buildings, which are more heterogeneous, and where the energy usage can be very dynamic, multivariate methods are often incorporated to try to control for effects for which comparison groups may not be adequate.

III.B.1. Institutional Retrofit Programs

A large fraction of the earliest efforts to evaluate commercial programs concentrated on institutional buildings that were targets of federal and regional conservation programs. These program evaluations included the federal Institutional Conservation Program (ICP) (87) and Bonneville Power Administration's Institutional Buildings Program (88, 89). They used a before-and-after analysis of billing records for participants alone. The results were less than completely convincing measurements of program impacts.

The largest problems were the lack of a comparison group to account for routine fluctuations in energy use and the small samples. Major changes in the operation and function of individual buildings, the lack of sufficient billing records, and the confounding effects of participation in multiple programs resulted in sample attrition.

Blachman's evaluation of BPA's program tried to use the participant buildings as their own controls, focusing on the variables like weather and energy prices in which change could be measured between pre- and post-program time periods (88). Unfortunately, with small samples and a large number of variables that couldn't be reliably quantified, there was little that the statistical approach could contribute. Davis' evaluation controlled for weather effects only (89). The Synectics approach for the federal program had a larger sample size, over 180 buildings nationwide, but did not include weather adjustments, partially because their sample depended heavily on measures to save oil, and the irregular schedule of oil deliveries was felt to add to the complexity of weather adjustment (87). [A more recent study reports progress in weather adjustment techniques for oil data (90).]

The U.S. Department of Energy initiated another institutional evaluation effort in 1985. Two approaches were eventually used. One approach used a second set of engineering estimates as an "evaluation" of program savings predictions made by program auditors (91). This research found that the predictions needed to be reduced many times, but that the two sets of engineering calculations were in general agreement. The second approach used historical billing records of hundreds of schools in Minnesota to compare trends in energy consumption over the entire period of program operation (92). Program-induced savings were identified despite a long-term trend toward efficiency in Minnesota schools. Schueler provides a more complete review of a decade of evaluation research on institutional building programs (93).

III.B.2 Commercial Retrofit Programs

The evaluations of non-institutional commercial programs have traditionally used billing analysis with a quasi-experimental design that has been the basis of most residential evaluations. Two of the earliest evaluations, both for California utilities, were exceptions. The evaluation of Pacific Gas and Electric's 1982 lighting conversion program used an MVR model to eliminate the non-program effects (94). It did not take non-participants into account. Again, with only a participant sample, an evaluation of SCE's program used a multi-equation model to account for both exogenous variables and the likelihood that the participants would have taken action on their own (37).

An early evaluation of a commercial lighting program in New England compared several techniques, but concentrated on a quasi-experimental design that the authors felt was the most useful of the designs attempted (95). Coates, in an evaluation of Bonneville's 1987 Commercial Incentives Pilot Program (CIPP) in the Seattle City Light service area, successfully employed a similar design with a small sample (N=37) (96). A full evaluation of the first two years of the CIPP across four utilities used the same basic evaluation design on a sample of over 200 buildings, with a comparison group of 190 buildings (97). The findings were consistent with the earlier results. Coates repeated the basic analysis of later program participants in Seattle, with stratified participant and comparison groups, and suggested that future analyses in the sector should move beyond the simple comparisons of participants and non-participants to a combination of multivariate methods (96).

In the Northeast, New England Electric System (NEES) employed quasi-experimental designs to estimate the impacts of its 1990 and 1991 Small Commercial/Industrial (C/I) Program and the 1990 Energy Initiative program for larger commercial and industrial consumers who installed only lighting measures (47, 98). The billing analysis was supplemented with a limited sample of end-use metering that supported the aggregate results (31). Commonwealth Electric in Massachusetts also used a quasi-experimental design for its small commercial program (99).

Combined approaches are being used in recent evaluations of commercial programs. In evaluating their 1991 large commercial program, NEES turned to a MVR model that adjusted for self-selection bias within a comparison-group design (100). The multivariate specification was an SAE "change model," in which the engineering estimates of savings from the auditors was a predictor variable, along with any energy-use factor that changed from the pre-year to the post-year, and the dependent variable was the change in consumption over the study period. For evaluations of Wisconsin Electric's C/I programs, a different approach has been developed, using both participant and non-participant comparison groups in a CDA analysis (39), and using end-use metered changes in consumption for a subsample of participants to "buttress" the precision of their engineering estimates of savings at the end-use level.

As has been the case in residential program evaluation, a sizable discrepancy between engineering estimates and actual savings often becomes evident when metered data are used. An example is the NEES evaluation of the Small C/I Program (31), in which engineering estimates were seen to overstate the measured savings for each type of lighting retrofit measure that was installed. Although no single explanation such as premature removal of the installed measures emerged, data from end-use metering and on-site surveys were found to be useful for exploring the reasons for the discrepancies.

III.B.3 New Commercial Construction Programs

If evaluation methodologies for commercial retrofit programs seem to be in their infancy, methods for the new construction programs must be in utero at this time. Unlike retrofit evaluations, with new construction programs there is no opportunity to compare before and after energy consumption. Nevertheless, evaluators did not start from scratch. As with retrofit programs, evaluators have had the chance to learn from residential evaluations. Yet, commercial buildings are dramatically more complicated. Not only are they a more heterogeneous group -- in part because they are defined so broadly -- but they have many large energy-consuming systems that are targets of programs. These systems include lighting, ventilation, refrigeration, and hot water, as well as heating and cooling. In addition, the buildings have large internal loads for computers and office equipment that are not the target of programs, but whose energy use interacts with that of the targeted end uses (101, 102).

Two approaches have been used to determine the net savings. The first uses building simulations. These allow the researcher to hold all factors equal except the efficiencies of the program measures involved. If the simulations are based on the design assumptions about the differences between the program and baseline buildings, the results are the equivalent of a priori engineering estimates. Post facto savings estimation involves re-simulating the building after it is built and occupied to capture both the changes in the actual construction from what was planned and changes in the assumed operation. (For details on this approach, see 103, 104.)

The second approach borrows from the experience of evaluating residential new construction. It uses the energy use intensities of participant buildings of a particular functional category, such as office buildings, and compares them in aggregate to similar non-participant buildings constructed at about the same time (105, 106).

Neither method has been frequently replicated. Both have problems: simulations rely heavily on the judgments of the individual modeler; not all conservation measures can be modeled; and the baseline building is subject to much interpretation (107, 108; see also 104). The use of whole-building energy indices may be an inaccurate measure if the program makes modifications to only one end use out of many within a building. In addition, the method will only work when there are large samples of similar buildings in each category within each region to use in the analysis, but there are generally too few buildings of some types, such as schools and hospitals, to allow for a sufficient sample. Multivariate approaches may help the comparison of energy intensities in the long run, but regressions too will require larger sample sizes. With most programs in this sub-sector still in their infancy, "verification" of the savings for regulators, to determine payments, relies on engineering estimates at this time (47, 53, 69).

III.C. Industrial Programs

By the late 1980s few evaluations of industrial conservation had been completed and hardly any were published (109) . Utility programs were very late getting started in this sector, and data on privately financed efficiency improvements tended to be proprietary. As utility programs developed, they either were combined with commercial programs, and emphasized lighting measures, or they were plant-specific efficiency improvements in industrial processes. The industrial retrofits in the combined programs are evaluated like commercial lighting, unless the portion of the whole facility load targeted for savings was too small to isolate with a billing analysis. In that case, site assessments are typically carried out to count the fixtures, determine the changes in installed lighting wattage, and estimate annual hours of operation.

Some industrial programs have involved case-study measurement approaches using site- and technology-specific measurements (110, 111). Sometimes these case studies can be combined to get a program-level picture of the savings impacts (48).

In the case of Northeast Utilities Energy Action Program, the process or major equipment retrofits were evaluated by detailed site assessments that resulted in a case-study estimate of what the actual savings probably were (112). True industrial programs will probably need to be measured on a site-specific basis (113). While many options are available to isolate the savings from confounding variables, such as production levels and product types (110), this approach cannot account for all confounding effects, especially the prior intent of the factory owner, and the possibility of free ridership. Site-specific methods will always have problems attributing the observed changes in consumption to the program.

No evaluations are known for programs aimed at new industrial facilities.

IV. IN CONCLUSION

The wide range of measurement methods, and the varied ways in which evaluators use these methods, make specific recommendations difficult about what evaluation approaches are best, especially for some of the newer types of DSM programs. More uniformity of methods, perhaps encouraged by evaluation standards and protocols, would in theory help evaluators learn from each other, so that this year's evaluations can become improvements on last year's. However, required use of methodologies that are not well developed or well understood is a "cart-before-the-horse" situation. This is particularly true for the estimation of end-use savings. DSM evaluation is such a new and rapidly growing field that adoption of more than minimal requirements at this time would be premature, and would surely squelch creative development of better methods. Coordinated development of methods, with research that attempts to shed light on reliability, potential bias and sources of inaccuracy, is needed outside the pressure of the regulatory arena, and with agreed-upon objectives for the methods.

The current use of disparate evaluation approaches makes meaningful comparisons across programs problematic. A recent review of commercial and industrial programs was comprehensive in its inclusion of data for over 200 programs

(114), but the absence of any systematic approach among utilities, as well as a reliance by almost all utilities on engineering estimates for the savings reported, renders impossible any meaningful comparison of the effectiveness of different DSM approaches. Probably because of a longer evaluation history, lessons learned in residential programs as a result of more uniform methods are encouraging.

One desirable feature of measurement methods is "transparency," so that inaccuracies and biases are not hidden within a complex methodology. A second, longer-range objective is standardization, wherein the same well tested methods are used across similar programs, whenever appropriate. A third objective is attribution, so that savings attributable to the program are separated from changes in consumption due to other factors. Finally, an essential criterion is "truth-seeking," to reveal what actually happened vs. what was expected, through a meter-based connection to reality.

It is clear from this review that no single measurement method satisfies all of these criteria for all types of DSM programs. Billing analysis, for example, is both transparent and tied to metered data. In addition, because of extensive validation research and applications to residential programs, billing analysis based on quasi-experimental design of participant and comparison groups has reached a high level of standardization. However, in DSM programs aimed a specific end uses, billing analysis does not satisfy the requirement of end-use attribution.

On the other hand, methods for estimating end-use savings generally make tradeoffs between a desire for attribution by end use, for which engineering estimates are inexpensive and flexible in their level of detail, and a desire for estimates of actual end-use savings, for which end-use sub-metered data are needed but for which the costs are too high to cover pre-program as well as post-program consumption, for comparison as well as participant buildings. An additional problem of end-use isolation is the potential neglect of interactive effects of one end use on another. Some methods are extremely complex, with disparate data sources and without well documented modeling techniques, and the resulting lack of transparency makes the results more difficult to interpret than for other approaches, less easily compared with results from other evaluations, and generally opaque to accuracy assessment.

The most promising evaluation approach is the use of multiple methods. When an end-use method is used, for example, the addition of a simple whole-building billing analysis provides an invaluable check on the accuracy of the end-use approach as well as an alternative view of program savings. If the whole-building and end-use savings estimates are consistent, the billing analysis will have bolstered the end-use estimation. If they are inconsistent, identification of the source of the discrepancy can provide valuable insight: lower savings from whole-building metered data than from the end-use estimates could signal inaccuracy in the latter, or confounding factors could have influenced whole-building consumption without affecting the end use(s) in question. In addition, billing analysis helps to answer the important question of whether end-use savings translate to savings at the whole-building level.

From past evaluations, it is clear that there are always surprises in the amount of energy actually saved. Identifying and understanding a shortfall in energy savings from a recent program could lead to enhanced savings in future programs. In this sense, the evaluation itself leads to savings. If, on the other hand, a shortfall is not identified, as it may not be if there is over-reliance on engineering estimates, the shortfall could be unnecessarily perpetrated through future programs.

The real truth of the effectiveness of DSM programs is whether expensive new power plants that the programs were designed to displace will indeed be averted over the long run. Eventually this truth will become evident, either as a lesson gleaned from DSM evaluations over the years, or as a "black eye" for Demand Side Management. Certainly, avoiding the need for new power plants is not the responsibility of evaluators. But avoiding a black eye <u>is</u> appropriately the evaluation community's responsibility. By working together to present an accurate and clear picture of the energy saved by DSM programs, evaluators can play a critical role in providing utilities and regulators with essential information for improving the effectiveness of planned programs.

Literature Cited

1. Hirst, E. 1990. Progress and potential in evaluating energy efficiency programs. Evaluation Review. 14(2):192-205

2. NARUC Conservation Committee. 1989. Resolution in Support of Incentives for Electric Utilities' Least-cost Planning: Convention Resolution No. 1, National Association of Regulatory Commissioners, Washington, DC

3. CLF 1991 Annual Report. 1992. Boston: Conservation Law Foundation, Inc.

4. Nadel, S. 1992. Utility demand-side management experience and potential -- a critical review. Ann. Rev. Energy Environ. 17:507-35

5. Keating, K. 1992. Persistence of savings. In Handbook of Evaluation of Utility DSM Programs, ed. E. Hirst, J. Reed. Report ORNL/CON-336, Oak Ridge National Laboratory, Oak Ridge, TN, pp. 89-99

6. Vine, E. 1992. Persistence of energy savings: What do we know and how can it be ensured? Proc. ACEEE 1992 Summer Study on Energy Efficiency in Buildings, pp. 7.257-64. Washington, DC: ACEEE

7. Kushler, M., Keating, K., Schlegel, J., Vine, E. 1992. The purpose, practice and profession of DSM evaluation: Current trends, future challenges. Proc. ACEEE 1992 Summer Study on Energy Efficiency in Buildings, pp. 7.1-25. Washington, DC: ACEEE

8. Saxonis, W. 1992. Free riders and other factors that affect net program impacts. In Handbook of Evaluation of Utility DSM Programs, ed. E. Hirst, J. Reed. Report ORNL/CON-336, Oak Ridge National Laboratory, Oak Ridge, TN, pp. 119-34

9. Fels, M., Reynolds, C. 1991. Toward standardizing the measurement of whole-building energy savings in DSM programs. Proc. 1991 Energy Program Evaluation Conference, pp. 75-85. Chicago: Argonne National Laboratory

10. Architectural Energy Corporation. 1992. Engineering Methods for Estimating the Impacts of Demand-Side Management Programs: Volume 1. Fundamentals of Engineering Simulations for Residential and Commercial End Uses. EPRI Report 3269-03, Electric Power Research Institute, Palo Alto, CA

11. RCG/Hagler Bailly, Inc. 1991. Impact Evaluation of Demand Side Programs: Volume 1. A Guide to Current Practice. EPRI Report CU-7179, Electric Power Research Institute, Palo Alto, CA

12. LBL. 1980. DOE-2 Reference Manual, Report LBL-8706 Rev. 2, Lawrence Berkeley Laboratory, Berkeley, CA

13. Fireovid, J., Misuriello, H. 1990. ASEAM-2.1 applications using parametric studies and multivariate regression techniques. Proc. 1990 ACEEE Summer Study on Energy Efficiency in Buildings, pp. 10.67-79. Washington, DC: ACEEE

14. Ecotope, Inc. 1984. SUNDAY 2.0, Ecotope, Inc., Seattle, WA

15. NJ BPU. 1990. Conservation Incentives Rulemaking. Docket No. EX90040304, Nov. 15:34. New Jersey Board of Public Utilities, Newark, NJ

16. Highest economy in furnace heating: proper temperatures, ventilation and coal consumption for different outside temperatures. 1906. The Metal Worker, Plumber and Steam Fitter. Nov 10:47-49

17. Bolton, R. P. 1911. Heating requirements in New York weather. Heating and Ventilation Magazine. VIII:1-10

18. Fels, M., ed. 1986. Special Issue Devoted to Measuring Energy Savings: The Scorekeeping Approach. Energy and Buildings. 9(1-2), 180 pp.

19. Schrader, T. F. 1978. A two-parameter model for assessing the determinants of residential space heating. MS Thesis. PU/CEES Report No. 69, Princeton University, Princeton, NJ

20. Goldberg, M. 1982. A geometrical approach to nondifferentiable regression models as related to methods for assessing residential energy conservation. Ph.D. Thesis. PU/CEES Report No. 142, Princeton University, Princeton, NJ

21. Campbell, D., Cook, T. 1979. Quasi-Experimentation: Design & Analysis Issues for Field Settings. Boston: Houghton Mifflin Company

22. Goldberg, M. 1986. A midwest low-income weatherization program seen through PRISM. Energy and Buildings. 9(1-2):37-44

23. Brown, M., Keating, K. ed. 1989. Special Issue Devoted to the Hood River Conservation Project. Energy and Buildings. 13(1), 91 pp.

24. Beschen, D., Brown, M. 1991. Evaluation Plan for the Weatherization Assistance Program. Report ORNL/TM-11668/V1, Oak Ridge National Laboratory, Oak Ridge, TN

25. Gregory, J. 1987. Ohio Home Weatherization Assistance Program Final Report, Office of Weatherization, Ohio Dept. of Development, Columbus, OH

26. Shen, L., Nelson, G., Dutt, G., Esposito, B., Fitzgerald, J., et al. 1990. The M200 enhanced low-income weatherization demonstration project. Proc. ACEEE 1990 Summer Study on Efficiency in Buildings, pp. 7.239-49. Washington, DC: ACEEE

27. Cohen, S., Goldman, C., Harris, J. 1991. Measured Energy Savings and Economics of Retrofitting Existing Single-Family Homes: An Update of the BECA-B Database. Report LBL-28147, Volumes I and II, Lawrence Berkeley Laboratory, Berkeley, CA

28. Reynolds, C., Komor, P., Fels, M. 1990. Using monthly billing data to find energy efficiency opportunities in small commercial buildings, Proc. ACEEE 1990 Summer Study on Energy Efficiency in Buildings, pp. 10.221-32. Washington, DC: ACEEE

29. Okstein, L. 1991. Measurement of energy savings in a utility performance contracting program. Sr. Thesis. PU/CEES Working Paper No. 123, Princeton University, Princeton, NJ

30. Spada, M., Peters, J., Goett, A. 1992. Evaluation of Boston Edison's small commercial and industrial retrofit program. Proc. ACEEE 1992 Summer Study on Energy Efficiency in Buildings, pp. 7.221-30. Washington, DC: ACEEE

31. Miller, M., Blake, W., Dagher, S., Schutte, J., Wright, R. 1992. Development of measured savings for a small commercial/industrial lighting program. Proc. ACEEE 1992 Summer Study on Energy Efficiency in Buildings, pp. 7.143-54. Washington, DC: ACEEE

32. Haberl, J. 1992. The use of a monthly whole-campus energy analysis for evaluating a third party energy service agreement. Proc. ACEEE 1992 Summer Study on Energy Efficiency in Buildings, pp. 3.95-110. Washington, DC: ACEEE

33. Pigg, S. 1992. Practical issues in using customer billing data for evaluating commercial, industrial, and large multifamily conservation programs. Proc. ACEEE 1992 Summer Study on Energy Efficiency in Buildings, pp. 3.213-23. Washington, DC: ACEEE

34. Ozog, M., Violette, D. 1990. Using billing data to estimate energy savings: specifications of energy savings models, self-selection and free-riders. Proc. ACEEE 1990 Summer Study on Energy Efficiency in Buildings, pp. 6.131-40. Washington, DC: ACEEE

35. Chi, M., Guinn, D. 1992. Billing analysis of low-income weatherization programs -- a case study: comparing PRISM to a multivariate regression technique. Proc. ACEEE 1992 Summer Study on Energy Efficiency in Buildings, pp. 4.29-39. Washington, DC: ACEEE

36. Moe, R., Buller, S., Miller, W. 1992. Billing analysis of HVAC customized rebates in the nonresidential sector. Proc. ACEEE 1992 Summer Study on Energy Efficiency in Buildings, pp. 7.155-63. Washington, DC: ACEEE

37. Train, K., Ignelzi, P., Kumm, M. 1985. Evaluation of a conservation program for commercial and industrial customers. Energy. 10(10):1079-88

38. Parti, M., Parti, C. 1980. The total and appliance-specific conditional demand for electricity in the household sector. Bell Journal of Economics. 11(1):309-21

39. Parti, M., Rogers, E., Sebald, A. 1991. Conditional Demand Analysis of commercial and industrial customers to determine the effects of a demand side program. Proc. 1991 Energy Program Evaluation Conference, pp. 323-28. Chicago: Argonne National Laboratory

40. Parti, M., Parti, C. 1992. Estimating the consumption impact of a residential conservation program. Proc. ACEEE 1992 Summer Study on Energy Efficiency in Buildings, pp. 4.177-81. Washington, DC: ACEEE

41. Sandusky, W., Pearson, E., eds. 1993. Special Issue Devoted to the End-Use Load and Consumer Assessment Program (ELCAP), Energy and Buildings, 19(3): 157-247

42. Eto, J., Akbari, H., Pratt, R., Braithwait, S. 1990. End-use load shape data application, estimation, and collection. Proc. ACEEE 1990 Summer Study on Energy Efficiency in Buildings, pp. 10.39-55. Washington, DC: ACEEE

43. Misuriello, H. 1990. Field monitoring of building energy performance. Proc. ACEEE 1990 Summer Study on Energy Efficiency in Buildings, pp. 10.77-93. Washington, DC: ACEEE

44. Hirst, E. 1987. Cooperation and Community Conservation: Final Report, Hood River Conservation Project. Report DOE/BP-11287-18, Bonneville Power Administration, Portland, OR

45. Dent, C., Denson, C., Duhn, F., Hadden, S., Abbott, R. 1991. 1991 Survey of Metering Equipment, Bonneville Power Administration, Portland, OR

46. Wright, R. 1989. Methods and Tools of Load Research - the MBSS System, RLW Analytics, Sonoma, CA

47. New England Electric System. 1992. 1991 DSM Performance Measurement Report, Massachusetts Electric Company, Westborough, MA

48. Riewer, S., Spanner, G. 1991. Performing impact evaluations in industrial retrofit: The Energy $avings Plan Program. Proc. 1991 Energy Program Evaluation Conference, pp. 485-93. Chicago: Argonne National Laboratory

49. Finleon, J. 1990. A method for developing end-use load shapes without end-use metering. Proc. ACEEE 1990 Summer Study on Energy Efficiency in Buildings, pp. 10.57-66. Washington, DC: ACEEE

50. Rohmund, I., McMenamins, S., Bogenrieder, P. 1992. Commercial load shape disaggregation studies. Proc. ACEEE 1992 Summer Study on Energy Efficiency in Buildings, pp. 3.251-62. Washington, DC: ACEEE

51. Akbari, H., Heinemeier, K., LeConiac, P., Flora, D. 1988. An algorithm to disaggregate commercial whole-building hourly electrical load into end uses. Proc. ACEEE 1988 Summer Study on Energy Efficiency in Buildings, pp. 10.13-26. Washington, DC: ACEEE

52. Northeast Utilities 1992. Results of the 15-Site Short Duration Monitoring Test, Northeast Utilities, Berlin, CT

53. Northeast Utilities. 1992. Determination of Energy Savings for Measures Installed in 1991, Northeast Utilities, Berlin, CT

54. Misuriello, H., Hopkins, M. 1992. Research Opportunities to Improve DSM Impact Estimates, California Institute for Energy Efficiency, Lawrence Berkeley Laboratory, Berkeley, CA

55. Violette, D. 1992. Analyzing data. In Handbook of Evaluation of Utility DSM Programs, ed. E. Hirst, J. Reed. Report ORNL/CON-336, Oak Ridge National Laboratory, Oak Ridge, TN, pp. 51-72

56. Xenergy. 1990. Demand Side Management Program Evaluation Scoping Study. Final Report ESEERCO Project EP90-34, Xenergy, Inc., Burlington, MA

57. Sebold, F., Fox, E. 1985. Realized savings from residential conservation activity. The Energy Journal. 6(2):73-88

58. Amalfi, J., Wright, R. 1991. Northeast Utilities' approach to C&LM program impact evaluation. Proc. 1991 Energy Program Evaluation Conference, pp. 191-99. Chicago: Argonne National Laboratory

59. Townsley, M., Wright, R. 1990. Measuring DSM impacts: end-use metering and the engineering calibration approach. Presented at 2nd Conference on End-Use Load Information and Its Role in DSM, Association of DSM Professionals, Irvine, CA

60. Hoffman, M., Parece, A. 1991. DSM evaluation methods: quantifying cost and accuracy (Identifying cost-effective alternatives to end-use metering). Proc. 1991 Energy Program Evaluation Conference pp. 314-22. Chicago: Argonne National Laboratory

61. Blasnik, M. and Lent, T. 1988. Quick savings estimates with GSTEAM: The GRASP Short Term Energy Analysis Method. Proc. ACEEE 1988 Summer Study on Energy Efficiency in Buildings, pp. 10.42-45. Washington, DC: ACEEE

62. Nadel, S. 1987. Timely evaluations: a discussion of methods used by the Massachusetts Audubon Society to evaluate fuel savings in under one year. Proc. 1987 Energy Program Evaluation Conference, Vol. 2, pp. 73-83. Chicago: Argonne National Laboratory

63. Kinney, L., Ucar, M. 1987. A cost-effective technique for determining by-measure conservation savings: several cold nights may be enough. Proc. 1987 Energy Program Evaluation Conference, Vol. 2, pp. 63-72. Chicago: Argonne National Laboratory

64. Hill, W., Blasnik, M., Greely, K., Randolph, J. 1992. Short-term methods for measuring residential energy savings: concerns and recommendations from comparisons with PRISM. Proc. ACEEE 1992 Summer Study on Energy Efficiency in Buildings, pp. 4.81-91. Washington, DC: ACEEE

65. Hirst, E. 1986. Electricity savings one, two, and three years after participation in the BPA Residential Weatherization Pilot Program. Energy and Buildings. 9(1-2):45-53

66. Hirst, E., Goeltz, R., White, D., Bronfman, B., Lerman, D., Keating, K. 1985. Evaluation of the BPA Residential Weatherization Program, Report ORNL/CON-180, Oak Ridge National Laboratory, Oak Ridge, TN

67. Brown, M., Brandis, P., Cody, B., Degens, P. 1991. Promoting energy-efficient home construction: the impacts of alternative policy instruments. Proc. 1991 Energy Program Evaluation Conference, pp. 547-54. Chicago: Argonne National Laboratory

68. Train, K. 1986. Qualitative Choice Analysis, Cambridge, MA: MIT Press

69. Boston Edison Company. 1992. First Annual 1991 DSM Program Reconciliation Report, Boston Edison, Boston, MA

70. Dutt, G., Lavine, M., Levi, B., Socolow, R. 1986. The Modular Retrofit Experiment: design, scorekeeping and evaluation. Energy and Buildings. 9(1-2):21-33

71. Hirst, E. and Hu, P. 1983. The Residential Conservation Service in Connecticut: Evaluation of the Conn Save Program. Report ORNL/CON-132, Oak Ridge National Laboratory, Oak Ridge, TN

72. Hirst, E., Goeltz, R., Thornsjo, M., Sundin, D. 1983. Evaluation of Home Energy Audit and Retrofit Load Programs in Minnesota: The Northern States Power Experience. Report ORNL/CON-136, Oak Ridge National Laboratory, Oak Ridge, TN

73. Berry, L., Brown, M., Wright, T., White, D. 1991. Experimental Plan for the Single-family Study: Weatherization Assistance Program, Report ORNL/TM-11668/V3, Oak Ridge National Laboratory, Oak Ridge, TN

74. Ternes, M., Levins, W., Brown, M. 1992. Experimental Plan for the Fuel-oil Study: Weatherization Assistance Program. Report ORNL/TM-11668/V2, Oak Ridge National Laboratory, Oak Ridge, TN

75. MacDonald, J., Brown, M., Ternes, M., Sharp, T. 1991. Experimental Plan for High Density Multifamily Study: Weatherization Assistance Program (draft). Report ORNL/TM-11668/V4, Oak Ridge National Laboratory, Oak Ridge, TN

76. Blasnik, M. 1989. Attrition bias in fuel savings evaluations of low-income energy conservation programs. Proc. 1989 Energy Program Evaluation Conference, pp. 211-17. Chicago: Argonne National Laboratory

77. Goldberg, M. 1990. PRISM for sparse data sets: a condensing lens. Proc. ACEEE 1990 Summer Study on Energy Efficiency in Buildings, pp. 10.91-99. Washington, DC: ACEEE

78. Hewett, M., Dunsworth, T., Miller, T., Koehler, M. 1986. Measured vs. predicted savings from single retrofits: a sample study. Energy and Buildings. 9(1-2):65-73

79. Ecker, L., Parti, M., Dion, M. 1989. A Comparison Study of Energy Savings Methodologies. Proc. 4th National Demand Side Management Conference, pp. 46.1-8, Cincinnati, OH

80. Nadel, S., Keating, K. 1991. Engineering estimates vs. impact evaluation results: How do they compare, and why? Proc. 1991 Energy Program Evaluation Conference, pp. 24-33. Chicago: Argonne National Laboratory

81. Granda, C. 1992. A statistically based impact evaluation of a direct install compact, fluorescent distribution program. Proc. ACEEE 1992 Summer Study on Energy Efficiency in Buildings, pp. 4.63-72. Washington, DC: ACEEE; and presentation at Conference

82. Violette, D., Buchanan, D. 1988. Evaluation of Wisconsin Public Service Corporation's Good Cents Home Program. Proc. ACEEE 1988 Summer Study on Energy Efficiency in Buildings, pp. 9.195-8. Washington, DC: ACEEE; and presentation at Conference

83. Starry, G., Michaelson, L. 1990. Evaluation of a new home construction program: Combining load research, billing data, and engineering estimates in a consolidated framework. Proc. ACEEE 1990 Summer Study on Energy Efficiency in Buildings, pp. 6.167-75. Washington, DC: ACEEE

84. Warner, K., Rufo, M. 1990. Comparisons of predicted and observed energy usage in Title-24 homes. Proc. ACEEE 1990 Summer Study on Energy Efficiency in Buildings, pp. 6.205-14. Washington, DC: ACEEE

85. Haeri, H. 1988. Energy Use in Homes Built to the Model Conservation Standards: A Follow-up of the Tacoma MCS Evaluation. Report IEAL/PO-31, Bonneville Power Administration, Portland, OR

86. Brandis, P., Cody, B., Brown, M., Haeri, H. 1990. Evaluating the impacts of the Model Conservation Standards in single family construction practices in the Northwest. Proc. ACEEE 1990 Summer Study on Energy Efficiency in Buildings, pp. 6.15-24. Washington, DC: ACEEE

87. Synectics Group. 1983. An Evaluation of the Institutional Conservation Program: Results of On-site Analysis: Final report. Report DOE/CS64571-1, U.S. Department of Energy, Washington, DC

88. Blachman, S. 1987. Institutional Buildings Program Impact Evaluation, Bonneville Power Administration, Portland, OR

89. Davis, B. 1987. The Institutional Buildings Program in Washington State: An Analysis of Program Impacts. Report Waoeng 87-34, Washington State Energy Office, Olympia, WA

90. Fels, M., Reynolds, C. 1992. Energy Analysis in New York City Multifamily Buildings: Making Good Use of Available Data, Report NYSERDA 93-3, New York State Energy and Research Development Authority, Albany, NY

91. Birdsall, B., Lebot, B., Kammerud, R., Hatfield, B. 1987. An Examination of Technical Audit Calculations for Schools and Hospitals. Report LBL-23537, Lawrence Berkeley Laboratory, Berkeley, CA

92. Vine, E., Hatfield, B., Lebot, B., Kammerud, R., Carroll, W. 1988. Energy Use in Minnesota Schools: Aggregate Performance and Perspectives on Energy Savings. Report LBL-24052, Lawrence Berkeley Laboratory, Berkeley, CA

93. Schueler, V. 1990. Measuring the impacts of energy efficiency measures in institutional buildings with billing data: A review of methodological issues. Proc. ACEEE 1990 Summer Study on Energy Efficiency in Buildings, pp. 6.155-66. Washington, DC: ACEEE

94. Minimax Corporation. 1984. Analysis of PGandE's Lighting Conversion Program: Energy Savings, Pacific Gas and Electric, San Francisco, CA

95. Nadel, S., Ticknor, M. 1989. Electricity savings from small C&I lighting retrofit program: Approaches and results. Proc. 1989 Energy Program Evaluation Conference, pp. 107-12. Chicago: Argonne National Laboratory

96. Coates, B. 1992. Longitudinal Evaluation of the Energy Savings in Commercial Incentives Pilot Program, Seattle City Light, Seattle, WA

97. Cambridge Systematics. 1990. Impact Evaluation of BPA's Commercial Incentive Pilot Program, Bonneville Power Administration, Portland, OR

98. New England Electric System. 1991. 1990 DSM Performance Measurement Report, Massachusetts Electric Company, Westborough, MA

99. MacDonald, M. 1992. Aggregate Impact Results: Commercial Direct Investment Program, Commonwealth Electric Companies, Wareham, MA

100. Davis, R., Lang, C., Ozog, M., Yokell, M. 1992. Impact Evaluation of the Energy Initiative Program, New England Power Service Company, Westborough, MA

101. Katz, G., Baylon, D., Heller, J., Heidell, J. 1989. Major Projects Rule: Phase II Evaluation, Seattle City Light, Seattle, WA

102. Norford, L., Rabl, A., Harris, J., Roturier, J. 1988. The sum of megabytes equals gigawatts: Energy consumption and efficiency of office PC's and related equipment. Proc. ACEEE 1988 Summer Study on Energy Efficiency in Buildings, pp. 3.181-96. Washington, DC: ACEEE

103. Harris, J., Diamond, R., deBuen, O., Hatcher, A., Nordman, B., Piette, M. 1990. Energy Edge Impact Evaluation: Findings and Recommendations from the Phase One Evaluation, Bonneville Power Administration, Portland, OR

104. Kaplan Engineering. 1991. Guidelines for Energy Simulation of Commercial Buildings, Bonneville Power Administration, Portland, OR

105. Harris, J., Diamond, R., deBuen, O., Nordman, B. 1990. Energy Edge Impact Evaluation: An Early Overview, Bonneville Power Administration, Portland, OR

106. Piette, M., Diamond, R., Nordman, B., de Buen, O. 1991. Memorandum on Additional Analyses of the Energy Edge Buildings, Bonneville Power Administration, Portland, OR

107. American Council of Engineering Consultants. 1988. The Energy Edge Project Building Design Process, Bonneville Power Administration, Portland, OR

108. American Council of Engineering Consultants. 1988. Energy Edge Process Evaluation, Bonneville Power Administration, Portland, OR

109. Spanner, G., Brown, D., Dixon, D., Garrett, B., Reilly, R., et al. 1988. Potential Techniques for Evaluating Industrial Conservation Programs under Bonneville's Energy $avings Plan, Bonneville Power Administration, Portland, OR

110. Harding, S., Gordon, F., Kennedy, M. 1992. Site Specific Verification Guidelines, Bonneville Power Administration, Portland, OR

111. BPA. 1992. Series of letter reports available for ten site-specific evaluations of industrial process projects, Bonneville Power Administration, Portland, OR

112. Northeast Utilities. 1992. 1991 Progress Report: Comprehensive On-Site Engineering Assessments and Ongoing Monitoring and Evaluation Activities, Appendix B to Ref. 53, Northeast Utilities, Berlin CT

113. Gervais, E., Haeri, H. 1991. Impact measurement of technology-diverse commercial and industrial programs: Challenges and options. Proc. 1991 Energy Program Evaluation Conference, pp. 494-97. Chicago: Argonne National Laboratory

114. Nadel, S. 1990. Lessons Learned: A Review of Utility Experience with Conservation and Load Management Programs for Commercial and Industrial Customers. Report NYSERDA 90-8, New York State Energy and Research Development Authority, Albany, NY