State of Florida

Public Service Commission

-M-E-M-O-R-A-N-D-U-M-

DATE: April 14, 1995

TO: Parties to IOU Electric Utility Conservation Program Dockets Nos. 941170-EG, 941171-EG, 941172-EG & 941173-EG

FROM: Joseph D. Jenkins, Director
Division of Electric and Gas JDJ

RE: Staff Communications

Although staff communications Rule 25-22.033 does not apply to PAA dockets, I am none-the-less hereby informing current and potential parties of a written communications received from LEAF. The written communication has voluminous attachments. Hence, I have enclosed LEAF's cover letter and note that the attachments are: (1) LEAF "Comments on Demand Side Management Plans of Florida's Investor Owned Utilities"; (2) "Handbook of Evaluation of Utility DSM Programs", by Eric Hirst and John Reed; and (3) "Monitoring and Evaluation Plan" (prepared for Iowa-Illinois Gas and Electric Company) by Xenergy.

Please contact Records and Reporting if you wish to see a copy of LEAF's attachments.

Also enclosed is a letter I received from Blalock and Associates.

JDJ/ng

enclosures

cc: Talbott
Bane
Records and Reporting w/LEAF att.
Palecki
Trapp
EAG Bureau Chiefs

DOCUMENT NUMBER-DATE
03824 APR 18 95
FPSC-RECORDS/REPORTING

BLALOCK & ASSOCIATES, INC.
ENGINEERING CONSULTANTS

763 FLAMINGO DR.
APOLLO BEACH, FL 33572
(813) 645-2901
FAX (813) 645-5664

BASE ENERGY CONSERVATION - NATURAL GAS INDUSTRY

NARRATIVE:

The Gas Industry and the Electric Industry are as different as baseball and football. About the only similarities are that both gas and electricity will flow through or down a metallic pipe, and both industries are regulated by governmental authorities. The rules which govern the activities of these two industries should be as different as the rules which govern play of baseball and football.

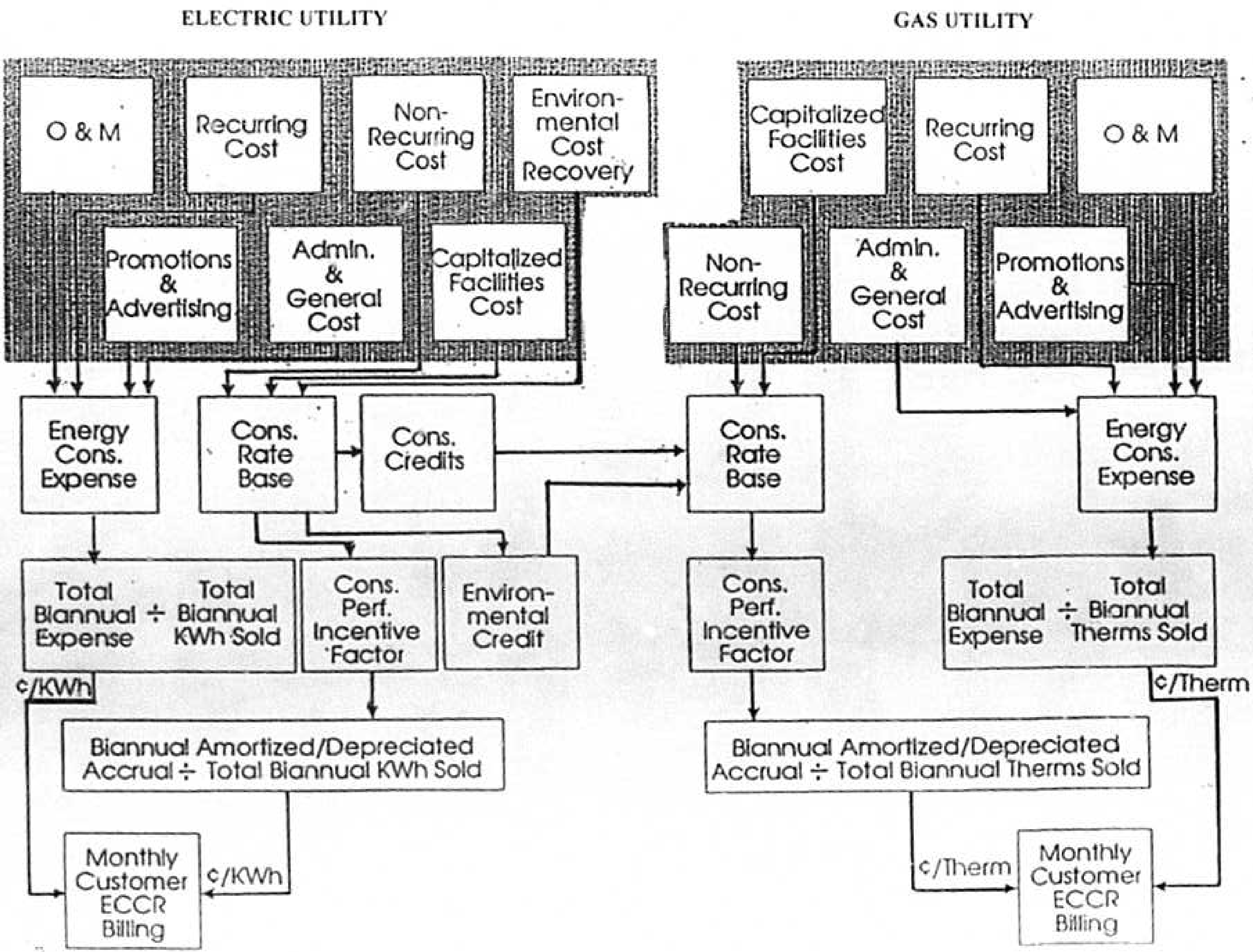
Historically, the rules which have been promulgated by governmental authorities governing the activity of the Gas Industry have paralleled and in many cases followed the rules which govern the activity of the Electric Industry. This is especially true with regard to "energy conservation" type programs and efforts which are suppose to conserve limited base energy resources and reduce and/or limit the increases in future Utility revenue requirements. The two industries diverge rather than converge or parallel each other in the "conservation" area because of the basic inherent economic and operating differences between the two industries.

The Electric Industry is characterized by high average incremental costs associated with incremental load additions, and high marginal energy costs associated with coincident system peaking conditions. The Gas Industry is characterized by low average incremental costs associated with incremental load additions, and low marginal fuel cost differentials associated with system peaking conditions. The two industries diverge in this respect. So, if conservation of scarce resources, least cost expansion options, and minimization of net present value of revenue requirements are the objectives of "energy conservation" programs and efforts, then the differences between the industries must be captured and used to the benefit of each industry's ratepayers in a manner which generates the highest net benefit to all ratepayers. This will require a complete overhaul of the "energy conservation" business and will require both industries to become energy conservation allies rather than energy supply competitors.

The cost to an Electric Utility to add one average residential customer is almost 3.5 times the cost to a Gas Utility to add one average residential customer. Indeed, the cost to an Electric Utility to add one KW of incremental load to their system cost almost (80%) as much as the total cost to add an entire residence to the Gas system. The cost to add one KW of incremental electric load to an Electric system cost over 30 times as much as adding an incremental appliance to a Gas system. The Electric Industry is far more "Capital Intensive" than the Gas Industry.

If cash-on-cash payback is considered when comparing the two industries, the differences are even more profound. The revenue flowing to an Electric Utility from either an average new residential customer or an incremental additional KW of load will pay back the capital costs to the Electric Utility in approximately five years. The average revenue flowing to a Gas Utility for the addition of a new residence requires almost ten years to be recovered; however, the Gas Utility only requires a few months (3 to 6) to recover the costs associated with incremental appliance additions. This is dependant upon the type of gas appliance.

The Electric Utility Industry is presently confronted with compliance with Phase I of the 1990 Clean Air Act Amendments. Most of the strategies which many of the electric utilities are implementing to gain compliance involve fuel switching (coal to gas) at the power plant (base load) level. The other most commonly used strategy employs "US Market-based Allowance Trading" which involves utilities who are in compliance selling "air pollution allowances" to utilities who need the allowances to gain compliance. The cost of compliance, regardless of the strategy and/or manner, will be recovered from the electric ratepayers through an "Environmental Cost Recovery" factor similar to CCR. The Compliance Plan Strategy must meet a least revenue requirement and/or a least net present value of revenue requirements test. The Gas Industry could be a major player in this process and increase retail gas sales and lower incremental operating costs by devising an "Environmental Credit" plan which in essence sells environmental credits to electric utilities who need the credits to gain compliance. In most cases this will positively impact the State economy by keeping all credits, allowances, and/or funds within the State.


ELECTRIC UTILITY
GAS UTILITY
O & M
Recurring Cost
Non-Recurring Cost
Environ-mental Cost Recovery
Promotions & Advertising
Admin. & General Cost
Capitalized Facilities Cost
Energy Cons. Expense
Cons. Rate Base
Cons. Credits
Total Biannual Expense ÷ Total Biannual KWh Sold
Cons. Perf. Incentive Factor
Environ-mental Credit
¢/KWh
Biannual Amortized/Depreciated Accrual ÷ Total Biannual KWh Sold
Monthly Customer ECCR Billing
¢/KWh
Capitalized Facilities Cost
Recurring Cost
O & M
Non-Recurring Cost
Admin. & General Cost
Promotions & Advertising
Cons. Rate Base
Energy Cons. Expense
Cons. Perf. Incentive Factor
Total Biannual Expense ÷ Total Biannual Therms Sold
¢/Therm
Biannual Amortized/Depreciated Accrual ÷ Total Biannual Therms Sold
¢/Therm
Monthly Customer ECCR Billing

SYSTEM/APPLIANCE EFFICIENCY
VERSUS
END-USE ENERGY RATES

A reciprocating engine, a combustion turbine, and a steam turbine all conform to the thermodynamic laws of nature. They all have base efficiencies on the order of 25% to 45% without considering any opportunities for waste heat recovery. Electric motors exhibit efficiencies between 85% and 95%. The end-use system efficiency (lighting, air conditioning, appliances, etc.) is not impacted by the energy source efficiency; however, the overall energy supply to energy consumption efficiency encompasses all efficiencies from the base energy source through the end-use.

Since low efficiency characterizes the power generation aspect of the Electric Utility energy supply equation and the same low efficiency characterizes the intermediate conversion of natural gas into mechanical force (compressor drive, generator drive, etc.) in the energy supply equation for the Gas Industry, the two energy suppliers compete with each other over a very narrow range of base energy conversion efficiencies and a very broad range of end-use efficiencies. This causes energy rates for both utility types (Gas & Electric) to favor certain end-uses and they are at opposite ends of the efficiency spectrum for each utility type. Gas rates favor low efficiency electric end-uses (water heating, lighting, small appliances,etc.) and electric rates favor high efficiency gas end-uses (air conditioning, refrigeration, etc.). The problem in Florida is that the Electric Industry is confronted with Summer and Winter weather sensitive electric demand and the Gas Industry is confronted only with a Winter supply problem and it is associated more with out of State supply problems. These differences cause the peak related marginal cost of energy supply to be almost completely non-coincident between the two industries. But, it offers great opportunity if the two industries cooperate in a plan to minimize the net present value of revenue requirements relative to common customers (Gas & Electric).

METHODOLOGIES FOR COST-EFFECTIVENESS DETERMINATION of BASE ENERGY CONSERVATION PROGRAMS

The Electric Industry has developed very data intensive and very mathematically rigorous methodologies for determining the cost-effectiveness of both energy conservation and system demand reducing programs. These methodologies and programs are founded in the avoided costs associated with delaying and/or negating future capital facilities expansion, and they are primarily directed at minimizing the net present value of future revenue requirements -- reducing rate escalation and conserving expensive and scarce base energy resources. There are basically two of these avoided cost methodologies and one derives from a perspective of a programs impact upon future rates (primarily a ratepayer test). The other derives from a perspective of a programs impact upon the various resources available to an Electric Utility and maximizing the net present value of the future benefits. In this methodology, minimizing future costs and revenue requirements is paramount (primarily a societal test). Capturing lost revenue and/or imbedded facilities costs are not a primary concern in the latter perspective.

Since system expansion for the Electric Industry is so capital intensive and since on-peak marginal fuel costs are so high relative to average and off-peak marginal fuel costs, these types of cost-effectiveness methodology lend themselves to an accurate evaluation or analysis of conservation and demand reducing measures. However, this type of methodology has no meaning if an attempt is made to directly extrapolate it to the Gas Industry. There are basically no future avoided costs associated with reducing gas sales. In fact, reducing incremental gas sales creates proportional incremental gas rate increases since the imbedded facilities cost is distributed over fewer gas sales.

An avoided cost methodology for determining the cost-effectiveness to electric ratepayers of gas alternatives to conventional electric conservation and demand reducing programs has been developed by the Gas Industry in Florida. This methodology is founded upon an electric utility's projected average avoided marginal fuel costs rather than its projected hourly avoided marginal fuel costs. Specific avoided generating units, specific generating unit in-service dates, and specific reserve capacity requirements are an integral part of the methodology. The primary objectives of the programs under consideration are "Base Energy Conservation" (saving gas, oil, and coal), and improving air quality.

A new methodology for determining the cost-effectiveness to gas ratepayers of these "Base Energy Conservation" programs needs to be developed by the Gas Industry in Florida. This methodology will be founded in the diminishing incremental operating costs which are associated with increasing gas sales, and the diminishing marginal fuel costs which are associated with "off-peak" (non-Winter) gas sales. The primary components of the reduced incremental operating costs which are a product of increasing gas sales are imbedded facilities costs distributed over increased gas sales, and a net revenue benefit which derives from the customer service charge. An "Off-peak Operating Cost Savings" element is also a benefit associated with the programs which facilitate off-peak as opposed to on-peak gas sales. The savings components which accrue as net benefits with this element are:

1) On-peak related operating cost such as regulator adjustments in Gate Stations, additional SCADA costs associated with Peak delivery of gas, and any manual by-passing and/or switching which operating personnel must do to adjust the system to Peak conditions.

2) Lost revenue (imbedded facilities cost) associated with cutting off interruptible customers during system capacity short falls (system capacity peaks).

3) PGA adjustments for On-peak deliveries by FGT and/or Suppliers.

4) Any differentials in gas purchase pricing which is related to On-peak versus Off-peak supply.

5) Any imbedded facilities costs which relate to supplying On-peak capacity (must be expressed as $/therm).

6) Any corporate overhead costs associated with purchasing and/or delivering On-peak versus Off-peak gas to the consumers.

7) Incremental system losses associated with On-peak conditions (higher pressures, higher flow rates, etc.).

8) Incremental variable costs associated with On-peak conditions (compressor operation, etc.).

A "Societal Benefit" (net BTU savings) is an integral part of these proposed "Base Energy Conservation" programs. This benefit is calculated simply by multiplying the weighed average heat rate of all the utility generating units (BTU/KWh) and the total electric energy (KWh) displaced and/or replaced, and then subtracting the gas energy necessary to displace and/or replace this electric energy. The ratio of the saved energy to the consumed energy could be used to create "conservation transfer payments" (incentives) between the Electric and Gas utility companies to promote this concept.

ACCOUNTING TREATMENT
of
COSTS, BENEFITS, INCENTIVE/REBATES, and TRANSFER PAYMENTS

A number of perceived disincentives presently exist as concerns the promotion of conservation and demand reducing measures by gas and electric utilities.

The electric utilities maintain that "customer provided capacity" ("non-traditional" power plant) denies them long-term investment opportunities and attendant earning opportunities, as well as short term energy sales (KWh) revenues. Further, they say that the magnitude of the electric utility's penalty is compounded by a strong commitment and successful implementation of these type measures. Florida Power Corporation and Tampa Electric Company favor "ratebase treatment" of some of the program costs such as equipment costs, installation costs, and one-time incentive payments to participating customers. They also advocate a "Conservation Performance Incentive Factor" (CPIF) to index performance to numeric goals and provide additional financial incentive, beyond allowed rate of return on ratebase, to encourage aggressive pursuit of program goals.

The gas utilities are confronted with disincentives of a completely different nature. The gas utilities who participate in Energy Conservation Cost Recovery (ECCR) were ordered by the FPSC to remove their Interruptible Rate class customers from ECCR. In many cases this removed a very significant portion of therm sales from the divisor used to calculate the ECCR factor, so the ECCR factor has become a disportionately high percentage of the remaining rate classes' bill calculations. Also, all of the expenses associated with "gas conservation" are treated as ECCR recoverable expenses and this magnifies the problem. Other disincentives are promotional costs such as advertising is generally disallowed, and one-time customer incentive payments create an undue impact upon the ECCR factor because they are collected over an arbitrarily short time frame (six months).

I offer the following as a means to implement appropriate treatment of costs associated with "Base Energy Conservation" programs which are determined to be cost-effective under the new cost-effectiveness methodology for these programs:

1) A "Conservation Ratebase" will be created and program costs such as capital facilities necessitated by approved programs, and one-time customer incentive payments associated with approved programs will be captured in the new ratebase.

2) A Conservation Performance Incentive Factor (CPIF) will be applied to the allowed rate-of-return on the Conservation Ratebase to encourage aggressive pursuit of numeric goals.

3) Transfer payments from Electric utilities to Gas utilities for electric ratepayer benefits associated with approved Gas Programs, credits for "Societal Benefits", and "Environmental Credit" will all diminish (write-down) the Gas Conservation Ratebase and will all be allowed charges to the Electric Conservation Ratebase with attendant DSM Goals crediting.

4) Approved program expenses which normally flow to ECCR will continue to flow to ECCR and will continue to be recovered bi-annually.

5) Coincident with the bi-annual ECCR review and true-up, approved charges and credits will be made to the Conservation Ratebase and a new per unit ($/KWh, $/therm, etc.) revenue requirement will be determined and collected through ECCR in addition to the other normally allowed ECCR collectibles.

6) All Conservation Ratebase charges including capitalized facilities costs and one-time incentive payments associated with approved programs will be amortized and/or depreciated in either a FPSC approved manner or in a manner generally accepted throughout the industry for the same type facilities.

ELECTRIC COST TO ADD ONE NEW RESIDENCE (4KW):

Generation Cost . . . $600/KW	$2400
Transmission & Distribution . . . $150/KW . .	$ 600
Meter & Service	$ 300
	$3300

ELECTRIC COST TO INCREMENTALLY ADD 1KW OF LOAD:

Generation Cost	$ 600
Transmission & Distribution	$ 150
	$ 750

GAS COST TO ADD ONE NEW RESIDENCE:

Meter Set & Service Cost	$ 840
Connection Cost (Customer Service) . . .	$ 24
Distribution Cost (Average).	$ 100
	$ 964

GAS COST TO ADD INCREMENTALLY AN APPLIANCE:

Customer Service Cost	$ 24

CONCLUSION

The array of "Conservation Programs" which are offered by most of the major Electric Utility companies in Florida primarily involve the deferral of electric energy into a different time frame - example, ON-peak time to OFF-peak time. This is generally accomplished by an electric demand control system which limits (reduces) the demand (rate of consumption) for electric energy. The total amount of electric energy for the end-use electric equipment being controlled is usually not significantly reduced, especially as compared to the demand reduction. The energy is certainly not displaced which implies that it was not made-up by another energy source and represents a lost energy (KWh) sale to the Electric Utility. An exception to this is an Electric Heatpump Conversion program where the Electric Utility offers an incentive to replace resistance heating with a heatpump. In this case the electric demand is reduced and the total energy sale over the useful life of the end-use appliance (heatpump) is also reduced. This represents a situation where an electric energy sale is lost for the life of the appliance, the electric demand is deferred for the life of the appliance, and the electric energy is displaced by the more efficient appliance for its entire useful life.

If the Gas Utility companies offer "Base Energy Conservation" programs which are cost-effective to both gas consumers and electric consumers, then, these programs can be structured to either defer electric energy and the rate of consumption of electric energy to another time frame, or totally displace (replace) electric energy and still meet the cost-effectiveness standards for both Gas and Electric utilities. An example is a Gas Utility subsidized program which accommodates an Electric Utility TOU rate and culminates in reduced energy operating costs to the electric consumer, deferred capital cost to the Electric Utility, and reduced base energy consumption to society. In this example some electric energy (KWh) is displaced and some electric demand is deferred; however, no electric demand is necessarily displaced or replaced in this example and it still meets the long term goals of both Gas and Electric utilities.

If the Gas Utility is constrained by conservation program specifications which call for electric demand and energy being totally displaced by gas end-use equipment, the cost-effective Gas program offerings will be severely limited. The primary reason for this is that electric rates and "Conservation" programs in Florida are driven or justified exclusively by the demand (KW) portion of the electric rates. This portion of the electric rate represents the capital cost for the Electric

Utility and the Gas Utility in many customer equipment applications cannot compete with just the remaining energy (KWh) portion of the rates. The Gas Utility "Base Energy Conservation Program" offerings should extend to both deferred and displaced electric demand and energy consumption in order to compliment similar Electric Utility Conservation programs, fill the natural customer participation voids which occur in all utility company conservation related program offerings, and offer maximum societal benefits.

LEAF

LEGAL ENVIRONMENTAL ASSISTANCE FOUNDATION

April 17, 1995

APR 17 1995

Mr. Joe Jenkins, Director
Division of Electric and Gas
Florida Public Service Commission
101 East Gaines Street
Tallahassee, FL

BY HAND DELIVERY

Re: Approval of Demand Side Management Plans for Florida's Investor-Owned Utilities
Docket Nos. 941170-EG through 941173-EG

Dear Joe:

Attached are LEAF's comments on the above referenced dockets. To help assure achievement of Commission-set conservation goals, LEAF urges you to recommend that the Commission:

- ***direct each utility to file and follow a monitoring and evaluation plan*** designed to cost-effectively gather information necessary to enforce Commission-set goals.
- make clear that ***goals cannot be met with savings from "free riders"***; and
- establish meaningful ***annual review*** of goals compliance.

These three actions would allow the Commission to reasonably assess conservation goals compliance and assure that conservation goals are met in the most cost effective way for Florida ratepayers.

Gathering the right information through monitoring and evaluation is critical to meeting the Commission's responsibility to see that ratepayers' money is reasonably invested. Without systematic measurement of DSM performance the Commission cannot reasonably determine whether DSM programs are providing the expected level of savings cost-effectively or take appropriate action to direct program modifications. Therefore, LEAF urges the Commission to direct the utilities to file monitoring and evaluation plans that meet the guidelines stated in the attached comments, Section III-D. For your reference, copies of a monitoring and evaluation plan that exemplifies many of these principles, and the Handbook of Evaluation of Utility DSM Programs, published by Oak Ridge National Laboratory are attached.

The Commission should also make clear that its conservation goals may not be met with savings from free riders, i.e., those who would have taken the identical energy conservation action ***without*** the DSM program. Counting free rider savings towards the goals would create peverse incentives to spend ratepayer funds on savings that would occur without utility intervention. Not counting free-rider savings appropriately directs utility efforts toward maximizing savings that would not otherwise occur--the reason for utility programs in the first place. Further, in calculating the cost-effective efficiency savings potential upon which the goals were based, the Commission used a load forecast that factored in the savings from free riders. Allowing goals to be met with free rider savings would effectively double count that savings--once in the load forecast and again in the DSM plan savings--degrading the goals to the detriment of utility customers.

Third, to assure that utilities reach their annual goals the Commission will need to verify goals compliance or shortfall and take appropriate action on an annual basis, for example as part of the standard ECCR process. Coupling goals compliance review with the annual ECCR prudence review of DSM spending offers significant advantages. Evidence regarding prudence could helpfully inform the Commission's assessment of appropriate remedies for goals shortfalls or rewards for goals achievements. Similarly, a review of savings achievements could indicate how effectively DSM budgets were expended and such findings could be germane to considerations of management prudence.

Thanks for your consideration. If you have any questions or wish to discuss these matters further, please let me know.

Sincerely,

Deb Swim

Deb Swim
Attorney
Energy Advocacy Project

cc: Parties of Record

LEAF
LEGAL ENVIRONMENTAL ASSISTANCE FOUNDATION

COMMENTS ON DEMAND SIDE MANAGEMENT PLANS OF FLORIDA'S INVESTOR OWNED UTILITIES

DOCKET NOS. 941170-EG through 941173-EG

April 17, 1995

A Public Interest Law Firm

1115 NORTH GADSDEN STREET • TALLAHASSEE, FLORIDA • 32303-6327 • 904-681-2591 • FAX 904-224-1275 *Recycled Paper*

CONTENTS

I. Introduction and Summary

In its final order in Docket Nos. 930548, 49, 50, 51-EG, the Florida Public Service Commission set annual energy and peak demand savings goals for the four major investor-owned utilities. The Commission established annual goals for the ten-year period 1994-2003, specified separately for the residential and commercial-industrial sectors. The Commission's decision required that goals be met in every year, and specified that sanctions would be applied if goals were not met.

The Commission-set goals are based on the utilities' estimates of the savings achievable with DSM measures and programs that pass the Rate Impact Measure Test ("RIM"). However, the Commission gave the utilities free rein to meet the goals through "whatever portfolios of programs they wish, including TRC programs" and encouraged utilities to "evaluate implementation" of RIM-failing/TRC-passing measures that offer large savings with nominal rate impact--inviting utilities to seek stockhold incentives and recovery of revenue losses for such measures. (Order No. PSC 94-1313-FOF-EG, p. 22)

The four utilities subject to the goals order--Florida Power & Light, Florida Power Corporation, Tampa Electric Company, and Gulf Power Company--have filed the first round of DSM plans since the order was issued. All four plans project sufficient savings from RIM-passing programs to meet the goals.

LEAF has reviewed the DSM plans and associated technical appendices. The utilities have not yet responded to our interrogatories concerning the detailed assumptions underlying the DSM plans. However, based on the sketchy information provided in the plans, we find that the filed program plans generally suffer from the following deficiencies:

- reliance on weak and needlessly expensive delivery strategies that create lost opportunities, promote cream skimming, and exacerbate free-ridership levels;
- lack of apparent consideration of the effect of free-ridership on overall savings levels (and thus contribution to goals) or program cost-effectiveness;
- failure to evaluate TRC options, either as less-costly substitute for RIM options in meeting goals or as cost-effective options for surpassing goals;
- absence of effective monitoring and evaluation ("M&E") protocols for tracking and verifying program performance and progress toward goals;

If utility M&E efforts are to provide useful information in time for the Commission's deliberations, it is essential that M&E plans be in place prior to the start of monitoring activities. The Commission should therefore direct utilities to file comprehensive M&E plans that prescribe appropriate M&E techniques for gathering the information that the Commission will need to enforce its goals directives.

We discuss in detail below the issues of program-performance verification, M&E planning, free-ridership, and goals enforcement as they relate to goals compliance. Based on the findings of our assessments of these issues, LEAF recommends that the Commission:

- direct each utilities to develop and file a comprehensive M&E plan consistent with the principles and guidelines stated below, and to collect and report the program-implementation data required to verify savings projections;
- explicitly state that goals cannot be met with free-rider savings, and require that achievements be reported both gross and net of free riders along with all assumptions and evaluation data regarding free-ridership;
- formally establish an annual reconciliation and enforcement process, either as an independent proceeding or as part of the standard ECCR process.

II. Verifying Goals Compliance

In its final order in the goals dockets, the Commission prescribed cumulative energy and peak-demand savings goals for the four major investor-owned utilities. The Commission ordered the utilities to meet these goals on an annual basis, but gave free rein to utilities to determine the best approach for complying with the Commission's directives:

> Utilities are free to file whatever portfolio of programs they wish, including TRC programs, in order to meet their goals. (Order No. PSC-94-1313-FOF-EG, p. 22)

In so doing, the Commission reserved the authority to review the performance of each utility's chosen course of action and to employ appropriate means for ensuring compliance with its directives:

> Any utility that does not achieve its goal shall be either penalized or have programs prescribed to it in a manner to be determined by this Commission on a case-by-case basis. (p. 22)

In essence, the Commission has defined its primary enforcement task as one of retrospectively reviewing actual program performance and reconciling performance with initial planning assumptions.[1] The Commission's review of actual spending and savings experience will be essential for validating utility claims regarding goals compliance in the previous year and, where necessary, for influencing future DSM efforts with prescribed programs or sanctions as appropriate. As part of this process, the Commission will need to address and rule on three issues:[2]

- Have utilities met the cumulative goals prescribed by the Commission for that year?
- Were DSM savings acquired cost-effectively and at minimum feasible expense to ratepayers?[3]
- Given a finding of noncompliance, what are the appropriate penalties and/or program revisions for effecting future compliance?

The Commission (and other parties) will need access to a body of program-implementation data sufficient to conduct a thorough compliance review. Since utility compliance filings will be relied on as the primary data source, it is critical that utilities provide not only overall spending and savings results, but the underlying program-performance data necessary to reconcile implementation experience with corresponding ex ante planning estimates.[4]

The Commission can use this reconciliation data in two ways. First, the data can be used to verify independently the utilities' findings regarding DSM-program cost-effectiveness, savings

---

[1]This is in addition to the Commission's review of program design proposals and retrospective prudence assessments.

[2]As discussed below, these issues can be addressed in conjunction with the Commission's regular prudence review of utility DSM spending.

[3]Although the Commission set goals based on utility estimates of RIM-passing savings potential, the final order in the goals docket does not explicitly require that utility efforts to meet the goals be cost-effective under the RIM or the TRC tests. However, the Commission presumably would require at a minimum that programs be cost-effective under the TRC, if not also cost-effective under the more restrictive RIM.

[4]The specific data required for a reconciliation analysis will vary by utility and program. We identify below the broad categories of data that would most likely need to be collected for any type of DSM program.

achievements, and goals compliance.

Second, if savings achievements fall short of the annual goal, the program-implementation data may reveal the cause of the shortfall and highlight appropriate responses by the Commission. For example, a utility may not have met its goal for the commercial-and-industrial sector due to a general economic downturn and consequent decreased savings potential from new building construction. In this case, the Commission might simply direct the utility to reallocate funding appropriated for new-construction programs to programs targeted to the retrofit market.

Alternatively, the shortfall may have been due to the utility's failure to raise rebates in the face of conclusive evidence that rebate levels were too low to promote program participation. If so, the Commission might choose to penalize the utility for its inaction and prescribe changes in program design to improve participation rates in the future.

As discussed in detail below, properly structured monitoring and evaluation of program results can provide the information the Commission needs to verify program achievements and improve program strategies. In turn, M&E efforts are most likely to provide useful information if guided by a cohesive and well-defined M&E plan. The M&E planning process serves to identify the overall purpose and goals of M&E activities, the program-implementation issues to be investigated, and the appropriate methods for gathering and evaluating program results. A properly designed M&E plan can ensure that M&E budgets are spent efficiently on securing the right information for answering the right questions.

Unfortunately, there is no evidence that the utilities have developed M&E plans or relied on such plans to select the M&E tasks described in their DSM plans. Monitoring and evaluation plans were not filed as part of the utilities' DSM plans. Moreover, the plans' cursory descriptions of proposed M&E activities give little indication of the overall purpose or goals of the tasks outlined or of the basis for relying on the evaluation techniques described therein.[5]

Without such M&E plans, the utilities' M&E efforts may proceed in a piecemeal fashion that results in serious gaps in the information collected and evaluated. If so, the Commission could lack critical data for determining actual program performance and overall goals compliance.

The need for complete and verifiable implementation data will be particularly acute for those utilities that are relying one or

---

[5]FPC's general discussion of the context and purposes of monitoring and evaluation is a notable exception.

two programs to provide the bulk of the compliance savings. For example, TECo expects that 80%-90% of its energy goals in the commercial-industrial sector will be met with savings from its Commercial Indoor Lighting Program. With so little diversification, there is significant risk of TECo falling short of its goal if the lighting program does not perform as expected.[6]

If utility M&E efforts are to provide useful information in time for the Commission's deliberations, it is essential that M&E plans be in place prior to the start of monitoring activities.[7] The Commission should therefore direct utilities to file a comprehensive M&E plan that delineates for each DSM program the issues to be investigated and the M&E techniques to be employed in addressing the issues of concern. We provide below general planning principles and guidelines to assist the utilities and the Commission in the timely development of comprehensive M&E plans.

III. Context and Purpose of Monitoring and Evaluation

Over the past decade, utilities have increasingly recognized and sought to acquire demand-side resources as part of a least-cost approach to meeting the energy service requirements of their customers. As DSM programs have become significant elements in utility resource portfolios, determining the effects of DSM programs has become increasingly important. Utilities, regulators, and customers want to know whether programs are providing the level of savings expected of them, whether the programs are cost-effective for the utility and for customers, and what can be done to increase participation, lower costs, and increase energy and demand savings.

Monitoring and Evaluation (M&E) is the process that has evolved to answer these questions. It is the "systematic measurement of the operation and performance" of DSM programs,[8] and a field of applied science where there are well-developed methods, standard practices, extensive literature, and a network of experienced technicians and

---

[6]If savings from this program are projected to contribute 80% of goals savings and actual savings are 10% less than projected, savings from the remaining programs would have to be 40% greater than expected to maintain goals compliance.

[7]Optimally, M&E planning should begin during the course of program design.

[10]Hirst, Eric, "Evaluating Demand-Side Management Programs," Electric Perspectives 14(6), 1990.

professional experts.[9]

As the magnitude of resources to be acquired by utilities through DSM increases, the importance of M&E also increases. Moreover, the rigor and the level of effort that should be applied to M&E becomes increasingly important and consequential. As noted by Hirst and Reed:

> Utilities should measure the performance of DSM programs with the same competence and diligence with which they monitor the performance of power plants. During the past several decades, utilities have developed detailed information on the construction costs, and the time, operation, maintenance and fuel costs; heat rate, availability factor, and capacity factor; the duration and causes of each outage; and fuel consumption, plant output, and emissions for each of their power plants....[10]

Monitoring and evaluation efforts need to be approached in an analogous manner, with a commensurate level of "plant engineering" (M&E planning), ongoing monitoring of the actual "plant output" (DSM program tracking and monitoring) and efforts to improve "plant efficiency" (DSM program refinement).

Monitoring and evaluation is generally recognized, both within the utility industry and in most regulatory contexts, to have three broad purposes: measurement; characterization and optimization.[11]

- Measurement has the purpose of quantifying the energy and demand impacts which are attributable to DSM programs, as well as the utility and societal costs of achieving those impacts.
- Characterization has the purpose of describing and analyzing the implementation of DSM programs, the markets addressed by programs and the market that remains, and identifying and

---

[9]See, generally, the proceedings of the bi-annual International Conference on Energy Program Evaluation, for the past ten years.

[10]Hirst, Eric, and J. Reed, Handbook of Evaluation of Utility DSM Programs, ORNL/CON-336, Oak Ridge National Laboratory, Oak Ridge, TN, 1991.

[11]Kushler, Martin, Ken Keating, Jeff Schlegel, and Edward Vine, "The Purpose, Practice, and Profession of DSM Evaluation: Current Trends, Future Challenges," Proceedings of the ACEEE 1992 Summer Study on Energy Efficiency in Buildings, Volume 7, American Council for an Energy Efficient Economy, Washington, DC, August 1992.

characterizing DSM resources that remain to be captured.

- Optimization has the purpose of refining program design and implementation based on program experience.

The measurement function of M&E will provide the quantitative data relied on to verify goals compliance. To meet the needs of the Commission, the measurement function should meet certain secondary objectives:[12]

* *Timeliness*--Measurement results should be available in a timely manner so that they may be considered by the Commission in the annual review of DSM portfolio achievement.

* *Proper Attribution of Impacts*--Measured energy and demand impacts should be disaggregated into savings attributable to the DSM program, including any spillover impacts, and those attributable to free riders or on-going adopters.

* *Objectivity*--The measurement results should be perceived as trustworthy; evaluations should be performed by parties with no stake in the results of evaluation.

* *Evidence*--Measurement results should be supported, within the constraints of cost, by evidence of the actual performance of installed DSM measures. Analysis and results should be documented in sufficient detail in order for them to be duplicated by others and allow independent review by the Commission.

* *Care and Technical Quality*--In order to be relied upon confidently, measurement and analysis should be conducted with the exercise of due professional care and consistent with the accepted standard practices and standards for technical quality that have been established for measurement and evaluation in DSM over the past 10-15 years.

* *Reasonable Accuracy and Precision*--Before programs are implemented, savings assumptions are merely estimates of probable impacts. Through the use of various measurement techniques, including metering, billing analysis, and refinement of engineering estimates based on program experience, M&E can be expected to substantially improve both

---

[12]Some of these objectives have been discussed in more depth in Butler, Susan, Dan Quigley and William Miller, "How Can We Assure Quality in Program Evaluation?" Proceedings of the Fifth International Energy Program Evaluation Conference, Chicago, IL, NEPEC, August 1991.

the accuracy and precision in the quantification of program impacts. While measurement can reasonably seek accuracy and precision on the order of ± 10 to 15%, the decision as to the appropriate levels to attain should be guided by analysis of the value of improved estimates relative to the M&E cost.

A. Impact Evaluation

The measurement functions of M&E are carried out through what are commonly referred to as impact evaluation activities, as differentiated from process evaluation activities. Process evaluation activities (discussed in the next section) fulfill the characterization and optimization purposes of evaluation. As discussed in the Handbook of Evaluation of Utility DSM Programs:

> Impact evaluations examine the effects of the program. They provide quantitative documentation of program benefits and costs. Impact evaluations measure program participation, participant acceptance of the recommended DSM measures and practices, performance of the DSM technologies promoted by the program, program energy and load reductions, and program costs.[13]

Impact evaluations compare what happened with the program to what would have happened if the program had not existed - with respect to customer's electrical requirements, customer costs, and utility costs. Such determinations are not as simple as examining the change in a program participants' energy use before and after their participation in a DSM program. First, it is necessary to determine the net savings rather that the total (or gross) energy savings:

> Total savings are the changes in annual electricity use and peak demand experienced by participants in the utility's program. Net savings are the portion of the total savings that can be directly attributed to the utility program.
>
> Net savings are the difference between total savings and the savings that participants would have achieved had the program not existed. Non-program savings reflect customer responses to changes in electricity and fossil fuel prices, changes in economic activity or personal income, introduction of new electricity-using technologies, and other non-program factors.[14]

---

[13]Hirst, Eric and J. Reed, Handbook of Evaluation of Utility DSM Programs, ORNL/CON-336, Oak Ridge National Laboratory, Oak Ridge, TN, 1991.

[14]Id.

Further adjustment of measured impacts is necessary to account for free riders, defined as those participants in a DSM program who "would have taken the identical energy conservation actions without the DSM program."[15] Free riders represent a cost to a utility's DSM programs, but provide no benefits (impacts). Accordingly the measured impacts of programs must be reduced to account for free riders and impact evaluations must include determination of, and adjustment for, free riders.

Conversely, many DSM programs can have the effect of achieving energy savings among some utility customers who are not formally program participants. These customers are often referred to as free drivers and their effect on program impacts referred to as spillover and market transformation. Free-drivership occurs when non-participating customers implement energy-efficiency measures either through a conscious awareness of the program or because of a program-induced change in the marketplace. An example of spillover would be where a customer purchases an efficient product based on information provided through program activities, but does not claim the rebate offered on the product. An example of market transformation would be where high participation in a new construction program transforms the market to the extent that higher levels of efficiency are implemented by non-participating builders for competitive reasons. While spillover and market transformation effects are difficult to quantify, there are established measurement techniques that can be used by utilities to account for these effects.

B. Process Evaluation

While the data needed to determine goals compliance is largely met through impact evaluation, utility M&E efforts should also be developed to address the characterization and optimization purposes of evaluation. If goals are not being met, it is the characterization and program optimization M&E results which will inform both regulators and the utility as to the probable reasons, the potential for achieving better results, and the program modifications that may be necessary so that goals can be achieved.

Even if there were 100% confidence that impact goals would be met, utilities would still be prudent to make use of the characterization and program-optimization functions of M&E. The feedback from characterization and program-optimization M&E activities can provide numerous benefits, including:

---

[15] Saxonis, William, "Free Riders and Other Factors that Affect Net Program Savings," in Handbook of Evaluation of Utility DSM Programs, ORNL/CON-336, Oak Ridge National Laboratory, Oak Ridge, TN, 1991

- refinement of estimates of market size, both with respect to potential participants and the DSM cost-effective resource potential of each participant;

- identification of opportunities for more comprehensive treatment of customers, with subsequently lower transaction costs;

- identification of successful program strategies that can used to improve other programs;

- improved understanding of market barriers to program participation and identification of more-effective strategies for overcoming those barriers;

- identification of barriers to persistence of savings and techniques to increase persistence;

- identification of measures with sub-optimal performance and methods to improve performance;

- better understanding of the effects of customer behavior on DSM-measure performance;

- better understanding of customer satisfaction with programs and opportunities for improvement

- identification of opportunities to improve program administration and program quality control.

These benefits have the potential for increasing program savings, reducing program costs, and/or improving customer service.

Characterization and optimization is achieved through process evaluations of DSM programs. Unlike monitoring and evaluation of supply options, which can rely largely on measurement techniques, DSM M&E necessarily must look to process evaluation for much of the necessary information to achieve evaluation goals:

> The major characteristic of DSM programs that distinguishes them from power plants is the distributed nature of DSM-program activities and effects. DSM programs are centrally planned, but are implemented by field staff, contractors and customers.[16]

Utilities have traditionally failed to devote adequate resources to process evaluation. A limited survey of utility DSM expenditures found that, on average, only about 30% of total M&E

---

[16]Bronfman, Ben and Jane Peters, "Process Evaluation of DSM Programs," Handbook of Evaluation of Utility DSM Programs, ORNL/CON-336, Oak Ridge National Laboratory, Oak Ridge, TN, 1991.

expenditures were budgeted for process evaluations.[17] However, the cost of a comprehensive process evaluation is likely to be comparable to that of a comprehensive impact evaluation.[18] The trend in recent years, particularly among utilities with more M&E experience, is to put more emphasis on process evaluations relative to impact evaluations.

C. The Need for M&E Planning

While evaluation is inherently a retrospective activity, it is essential that M&E be systematically planned in advance of its implementation and that the results of the planning process be documented in an M&E plan. Without adequate and timely planning, evaluation will likely be inefficient and unable to achieve its goals. The types of data that are required for program evaluation cannot be anticipated without going through a planning process that identifies the data necessary to answer program-specific evaluation questions. Equally important, a wide range of both impact- and process-evaluation techniques and experimental designs require the collection of data <u>during</u>, and in some cases <u>before</u>, the period of program implementation. Thus, M&E should also be planned in advance of the program implementation.

Ideally, M&E planning will be an integral part of the program-design phase of DSM program development. Just as program design influences evaluation requirements, contemporaneous evaluation design can influence the nature and timing of planned program activity. For example, delivery mechanisms and schedules may be modified to take advantage of data collection opportunities.[19] If evaluation planning cannot be conducted simultaneously with program design, it should be completed prior to program implementation which is the subject of evaluation.

The benefits of planning prior to program implementation include:

- Establishment of hypotheses to be tested, which in turn define data needs and analysis methods.
- Assurance that early evaluation activities can provide prompt

---

[17] Wirtshafter, Robert, "Establishing Priorities for Future Evaluation Efforts," Proceedings of the Fifth International Energy Program Evaluation Conference, Chicago, IL, NEPEC, August 1991.

[18] Hicks, Elizabeth, "Planning Evaluations," Handbook of Evaluation of Utility DSM Programs, ORNL/CON-336, Oak Ridge National Laboratory, Oak Ridge, TN, 1991.

[19] Hicks, Elizabeth, op. cit.

feedback to program implementers, enabling early identification of problems and timely corrective action.

- Assurance that program designs support the data-collection needs.

- Full use of data collection opportunities during implementation, ensuring that all necessary data will be collected, and reducing costs for data collection and analysis.

- Enhanced data quality and consistency, and establishment of precision requirements and sample designs.

- Facilitation of efficient design and development of a DSM tracking system to store and maintain data.

Without a detailed M&E plan that establishes research hypotheses, experimental designs, and data needs and analysis methods, substantial opportunities for efficient data collection and early implementation feedback will almost certainly be lost. With proper planning, utilities can ensure that the program-design supports the necessary data needs, and that data is collected efficiently and effectively.

The results of the M&E planning process should be reported in an M&E plan for two reasons. First, the evaluation plan plays an important management role as a reference document for the utility and its contractors to implement evaluation activities. Second, M&E plans should be available for critical review by the Commission at the time of utility program plan filing. Without these plans, the Commission will not be able to determine whether the utility has developed effective strategies for collecting and evaluating the data needed to assess program performance and verify goals compliance.

Data collection should not be confused with evaluation. Collecting appropriate data is necessary, but not sufficient, for proper evaluation. Many utility evaluations suffer from inconsistent or missing data, either because the data were collected with inconsistent methods, or defined inconsistently. This is often the result of poor planning, with data being collected at different times, from different participant or non-participant samples, with different survey instruments, or stored in different databases designed without a clear understanding of the ultimate uses of the data. Proper planning establishes what data are needed, how they should be defined and used, from whom they should be collected, how and by whom they will be collected, and where and in what fashion they will be stored and kept up-to-date.

At best, the consequences of not planning will be that substantial cost and effort is expended to collect supplemental data after the fact, to merge databases, and to perform quality-control and data-cleaning exercises. At worst, the failure to plan will result in the introduction of significant bias into the analysis results, or the inability to properly analyze specific issues, such as free ridership, persistence, or snap-back. The potential costs of these deficiencies are much greater than the cost of data collection alone. These deficiencies may result in erroneous conclusions regarding whether goals are being met. This in turn can lead utilities to continue to implement programs that are not cost effective, miss highly cost effective DSM opportunities from program enhancements, or collect inappropriate incentives from ratepayers.

D. M&E Planning Guidelines

This section provides general guidelines for developing and implementing effective M&E plans. These guidelines reflect a general consensus among leading practitioners in the M&E industry about the best approaches to M&E planning.

As first summarized by Hicks[20] and usefully elaborated on by McRae, et. al.,[21] the M&E process can be described as having six steps:

1. Identify evaluation objectives and the program-specific research questions to be answered by M&E.

2. Identify evaluation approaches capable of answering the research questions and providing information suitable for decision making.

3. Assess alternative experimental designs, techniques, and analytical methods, including their data requirements.

4. Estimate the value of information and use this to establish an appropriate budget for M&E activities.

5. Conduct the evaluation.

6. Communicate the results.

Reflecting this process, an M&E plan should comprise the

---

[20]Hicks, Elizabeth, op. cit.

[21]McRae, Marjorie, Henneberger, and Hanser, "Now that We've Got Their Attention: Guidelines for Producing Useful and Used Evaluations," Proceedings of the ACEEE 1992 Summer Study on Energy Efficiency in Buildings, Volume 7, American Council for an Energy Efficient Economy, Washington, DC, August 1992.

following:

- A clear statement of the purposes of DSM M&E for the subject utility, reflecting the interests of all stakeholders.

- A brief description of each DSM program being evaluated, including history and annual goals for participation, measure penetration, and impact goals.

- A detailed list of the hypotheses to be tested, and how the results will be used to improve program design and implementation.

- A description of the research methods, experimental designs, and analytical methods to be applied (with identification of approaches to be used in addressing each of the identified hypotheses).

- Identification of the data needs and a data collection plan.

- Estimates of the accuracy and precision of the selected M&E approaches to be implemented, including sample designs where applicable.

- Detailed implementation plans and schedules for each type of evaluation.

- Detailed evaluation budgets, including support for appropriateness of the proposed budget relative to the value of the information procured through evaluation.

- Identification of the deliverable products, timetable for completion, and deadlines for when results are needed for decision making (e.g., regulatory reviews, budgeting cycles, input to IRP).

When reviewing M&E plans, utility staff members and regulators should also look for the following attributes:

- M&E plans should explicitly identify how the findings will be incorporated into future program design and delivery decisions, and in utility planning. Without these issues clearly articulated, it is likely that research and data collection will suffer from a lack of direction.

- M&E plans should establish a detailed schedule, budget and resources. The schedule should integrate the evaluation with program design and implementation activities. Data must be collected in time to support each phase of the analysis, and results must be available in time to support future program decisions.

- The evaluation budget and staffing plan should ensure that upper level managers are committed to the evaluation, that appropriate staff members and contractors will be available in a timely fashion to complete M&E tasks, and the planned expenditures are appropriate given the program design, projected impacts, and future importance of the program to the utility's overall DSM plans.

M&E budgets typically range from about 5% to 10% of program implementation costs. In a survey of utility DSM expenditures, the range of expenditures observed varied from 3% to almost 30%, but most utilities appear to be guided by a 5%-to-10% rule of thumb.[22] However, as noted by Wirtshafter and numerous other evaluation researchers, a single rule-of-thumb is not appropriate; ultimately, budget levels should be reconciled with the value of the information procured.[23] Pilot programs, or programs in early phases of development may warrant relatively large expenditures on M&E (often 30% or more). This is particularly true if they are testing different delivery or incentive mechanisms, or are being analyzed for replication on a larger scale. On the other hand, large, mature full scale programs may require relatively small expenditures on evaluation (as a percent of implementation cost). This is particularly true of large rebate programs, where costs primarily represent incentives. Another factor is the absolute size of the program. Because some M&E costs are relatively fixed regardless of program size, M&E costs as a percentage of total program costs will be larger for smaller programs.

The M&E goals and objectives may also dictate appropriate levels of expenditures. For example, while a large and mature full-scale program may have less need for extensive process evaluation, it may represent a large portion of the utility's projected DSM impacts. Therefore, a relatively large portion of impact-evaluation funds should be spent to ensure accurate impact estimations. Ultimately, impact-evaluation expenditures should be allocated based on projected program impacts, the projected uncertainty in

---

[22]McRae, Marjorie, Henneberger, and Hanser, "Now that We've Got Their Attention: Guidelines for Producing Useful and Used Evaluations," Proceedings of the ACEEE 1992 Summer Study on Energy Efficiency in Buildings, Volume 7, American Council for an Energy Efficient Economy, Washington, DC, August 1992.

[23]Hicks, op. cit. Semak, Michael, Robert Uhlaner and Bruce Smith, "Building Reliable DSM Resources with Program Evaluation," Proceedings of the ACEEE 1994 Summer Study on Energy Efficiency in Buildings, Volume 8, American Council for an Energy Efficient Economy, Washington, DC, August 1994. Hummel, Philip, "Resource Allocation and DSM Program Evaluation Planning," Proceedings of the Sixth International Energy Program Evaluation Conference, Chicago, IL, NEPEC, August 1993.

current impact estimates (from simple engineering analyses), and the projected marginal costs of precision improvements.[24]

An M&E plan should identify precision requirements, and how these will be met. Precision refers to the range of variance around an estimate given a probability of certainty. In other words, one might estimate impacts with 90% confidence that the actual impacts are within plus or minus 20% of the estimate (20% precision at 90% confidence). Greater levels of precision generally require either larger samples or more robust techniques, and typically will involve greater cost. Thus, the plan should include a discussion of cost versus precision t rade-offs, and the utility's rationale for the selected level of precision. In addition, a sampling plan should be included that supports the established level of precision, and that is reasonable given the hypotheses to be tested, data collection limitations, and time and resource constraints. Sample designs should be sufficiently detailed to ensure their efficiency.

M&E plans should identify the research methods to be used. These should be consistent with, and a logical extension of, the M&E goals and objectives, and fully support the hypotheses to be tested. Discussion should include the rationale for the method(s) selected, the data needed to support each approach, and how, when and by whom it will be implemented. The plan should also identify the information resulting from the analysis and how it will be used to support impact estimates or program decisions.

Perhaps one of the most critical functions of an M&E plan is full disclosure of the data needs to support M&E, and a data collection plan. Some data may need to be collected prior to program implementation (e.g., baseline data). Therefore, it is critical that data needs are fully identified up-front. In addition, a proper procedure for data storage and handling is necessary for effective on-going monitoring. The plan should identify exactly how data will be collected, including what samples will be used, what types of survey instruments and methods will be used, who will collect data, when it will be collected, and the procedure for data input, storage, handling and quality control. In addition, the plan should explicitly identify how the data will be used in the analysis. This will enable reviewers to ensure that all necessary data will be collected prior to analysis.

The following is an example of the types of data that should typically be collected to support impact and process evaluations:

---

[24]For more information on allocation of evaluation costs, see Horowitz, M., DSM Impact Evaluation Planning: Standardization Versus Diversification, forthcoming paper in the National Conference on DSM Evaluation proceedings, Chicago, IL, August 1995.

* *Direct measurement data.* These will typically support impact evaluations and may include whole premise, end-use, operation (run-time), equipment-metering, and customer-billing data.

* *Engineering data.* These include facility-audit data; baseline equipment, operating and behavioral data, weather data, and data on installed measures.

* *Statistical data.* These data typically support statistical impact estimates and are also used in process evaluations. They may include such things as customer and demographic data, segmentation data, regional data, and economic variables.

* *Market data.* These include information to support process and market evaluations, including customer and trade-ally preferences, decision-making factors, behavior, satisfaction with program services and measures, and equipment-sales and distribution data.

Particular attention should be paid to the methods and timing of data collection. For example, engineering data typically must be collected on-site, by a trained auditor; engineering equipment and operating data collected through mail or phone surveys will typically be incomplete or inaccurate. A data-collection plan will ensure that lost data collection opportunities are not created. For example, if on-site visits are included as part of program delivery, data collection should be done at the same time. Often short-term metering equipment can be installed during pre-inspections, read at measure treatment, and then collected during post- inspections. Other lost data-collection opportunities occur when utilities fail to coordinate with other utilities to implement joint data collection or analyses. These opportunities might involve neighboring utilities collecting trade-ally data on current market trends to assist in baseline and free-ridership estimation.

E. Critical Issues and Approaches

Monitoring and evaluation of DSM programs is an evolving practice, and can often involve numerous and sophisticated techniques. A detailed discussion of these techniques is beyond the scope of these comment.[25] However, the following addresses some general guidelines concerning free-ridership estimation, joint-utility activities, and overall impact estimation. Free-ridership and overall impact estimation are discussed because of their

---

[25]For more detailed discussions of M&E research techniques, see Hirst, E & J. Reed, et. al., Evaluation of DSM Programs, Oak Ridge National Laboratories, Oak Ridge, TN, 1991; and Violette, D., et. al., Impact Evaluation of Demand-Side Management Programs, Electric Power Research Institute, Palo Alto, CA, 1991.

critical importance to establishing whether utilities are meeting their goals, and because of the inherent biases and inaccuracies in many estimation approaches. Joint-utility activities are singled out because they offer substantial cost saving benefits and are often neglected by individual utilities, but can be effectively promoted by regulators.

F. Free Ridership

A primary goal of M&E is to accurately quantify costs and benefits attributable to a DSM program. Initial program trac king system estimates typically represent gross impacts, based on the savings realized from all participants. To properly determine the impact of a DSM program, net impacts should be calculated by subtracting out free riders. Free riders are program participants who would have taken the same action in the absence of the program. If a customer's action is identical to what he or she would have taken without the program, they are a pure free rider, and the benefits realized by their action are not properly attributable to the program. Typically, a participant may be only a partial free rider. The customer may have been influenced by the program to implement additional measures, an incrementally higher efficiency measure, or to implement the measure sooner than they would have done without the program. In this case, only a fraction of the customer's savings are attributable to the program.

The extent of free-ridership is likely to be significantly influenced by program planning and design features. This can happen in a variety of ways. For example, if planners underestimate the baseline equipment efficiency, a rebate program may offer rebates for measures that already enjoy significant market saturation. In fact, some utilities have previously offered rebates for equipment with efficiencies that do not exceed code requirements, virtually guaranteeing high free ridership. A perhaps less obvious, but more common, program design that may result in high free ridership is when utility incentives are so low that they do little to influence customer behavior. Under this situation, only customers already planning to install a measure are likely to participate. Because program-design features can have a significant impact on free ridership, early and accurate free rider estimates are important, not only for ex post determinations of savings achievements, but to provide early feedback to support program modifications.

Unfortunately, while free-ridership estimation is a critical component of M&E, it is still an inexact science.[26] Previous free-

---

[26]See, for example, Lui, D. & J. Fang, "Issues in Free Rider Estimations," Proceedings of the 1990 ACEEE Summer Study on Energy Efficiency in Buildings, Vol. 6, 1990; Kreitler, V., "Market Issues in Free Rider Estimation," Proceedings of the Electric Utility Marketing Research Council EUMRC/EPRI Conference, EPRI CU-

rider estimates have varied widely, and have often suffered from significant bias.[27] Four primary methods are typically used to address free ridership:

- participant surveys;
- measure sales data and baseline practices;
- comparison groups;
- statistical modeling.

The most common method for analyzing free riders is to survey (either by phone or mail) program participants. This has the advantage that it is relatively inexpensive and relies on primary data about the population in question (i.e., participants). Unfortunately, it is often difficult to obtain accurate answers from customers about hypothetical actions. If one simply asks a customer whether they would have done the same thing without the program, results tend to be biased downward. This is because customers will tend to answer the way they believe the surveyor desires.

Methods to overcome bias include using a series of questions that, taken together, allow the analyst to make judgments about the customer's likely behavior absent the program. For example, one might ask a customer to describe what plans they had prior to knowing about a program, and follow up with questions about the specific details of these plans and how much they would have paid for specific measures. Often participants who said that they planned to install efficient equipment indicated, in response to questions, that they did not have any specific plans, or would not have spent the necessary funds to follow through.

Even with well-designed surveys, customers may not know or remember what they would have done absent the program. As a result, participant surveys often are combined with other methods. One approach is to analyze equipment sales and distribution data. This approach may entail identifying pre- and post-program sales or shipments of equipment, both within the service territory and in a

---

7010, prepared by SRC, Bala Cynwyd, PA, 1990; and Saxonis, W. "Free Riders and Other Factors That Affect Net Program Impacts," Evaluation of Utility DSM Programs, ed. Hirst, E. & J. Reed, Oak Ridge National Laboratory, Oak Ridge, TN, 1991.

[27]Saxonis, W., "Free Riders and Other Factors That Affect Net Program Impacts," Evaluation of Utility DSM Programs, ed. Hirst, E. & J. Reed, Oak Ridge National Laboratory, Oak Ridge, TN, 1991, p. 125, Table 24, "Example of Free Rider Estimates" shows a wide range of free rider estimates by program and technology type.

shipments of equipment, both within the service territory and in a comparison area without a program. Possible sources of data include surveys of vendors, manufacturers, and contractors, or secondary data collected by industry trade associations. One drawback to this approach is that sales data can not be disaggregated into participants and non-participants. Therefore, an increase in equipment sales does not necessarily mean the increase was due to program participation. It is also difficult to determine the underlying causes of market changes. For example, a similar increase in equipment sales both in the program territory and a neighboring region may indicate high free-ridership (the baseline practices shifted), or high free-drivership (the increase in sales outside the service territory was a direct result of spillover from the program).

Finally, statistical approaches, such as discrete choice models, can be used to model customer behavior (of participants and non-participants) and determine the importance of different factors in their decision-making, including participation in the program. These approaches still require extensive participant and non-participant survey data, and are often used in conjunction with other analysis methods.

G. Impact-Estimation Techniques

Impact estimation techniques typically involve one or more of the following methods:

- engineering estimates,
- direct metering,
- statistical billing analysis.

These methods should be combined to both improve precision and reduce costs. Engineering estimates, by themselves, are often inaccurate and biased. They rely on a deterministic approach to impact estimation that makes assumptions about customer behavior and equipment operation, often without primary data. Metering can provide extremely precise measurements of actual energy consumption and demand, but is costly. In addition, both engineering and metering approaches suffer from the inability to capture such behavioral effects as snapback and customer response to price changes and underlying economic factors.

Unlike engineering and metering methods, statistical billing analysis (SBA) can capture economic and behavioral effects by comparison with an appropriate control group. In addition, SBA is often considerably less expensive than metering. However, SBA may suffer from bias resulting from self selection or other factors, and often provides estimates with unacceptably large confidence intervals. This may be particularly true if utility incentive payments are based on impact estimates.

One way to improve SBA precision and reduce bias is by incorporating engineering and metering data into SBA model specifications. For example, a small metering sample can be used to determine bias in engineering estimates on a larger sample. These estimates can then be improved using a ratio adjustment. In turn, the adjusted engineering estimates can be used as a variable in a statistically adjusted engineering model to reduce the variance of the SBA performed on a relatively large sample. These hybrid techniques should be considered by utilities in their M&E plans to maximize impact estimation precision at a minimal cost. Some utilities have spent large sums on metering, while ignoring other methods. The result is inordinately large M&E expenditures, and little information about true net impacts, because of an inability to identify behavioral effects and free riders. Other utilities have relied solely on SBA, ignoring valuable engineering or metering data available to them that could significantly improve estimates.

H. Joint Utility Methods

Florida utilities should consider combining research efforts on issues of common concern. Often, competing utilities will reinvent the wheel by doing separate analyses, when a single co-funded analysis would result in lower cost to all ratepayers and more accurate information. Examples of joint projects might include non-participant surveys, analysis of baseline practices, trends in equipment sales and saturation, engineering studies, and estimation of spillover effects. Utilities in New England and Vermont have recently used this collaborative approach.[28]

IV. Accounting for Free-Ridership

The magnitude of savings and utility spending attributable to free riders could prove to be a critical consideration in the Commission's verification of program achievements. If actual free-ridership exceeds initial planning estimates relied on by utilities to set the pace of program implementation, actual program-savings achievements could fall short of the goals set by the Commission. For example, anticipating an average free-ridership rate of 10%, a utility might set its DSM budget for the next year so that sufficient savings from non-free riders are acquired to meet its annual goal. If actual free-ridership amounts to 20%, non-free-

---

[28] In New England, EUA, BECo, Commonwealth Electric, NU and NEES jointly implemented spillover and residential-new-construction baseline studies. A project in Vermont also looked at new-construction baseline practices.

rider savings will fall short of the goal by 11%, all else equal.[29]

Free-ridership reconciliation will be an issue because free-rider savings should not be allowed to contribute to meeting a utility's goals. If free-rider savings were counted towards goals, then the difference between estimated and actual free-ridership will have no effect on goals accounting. However, because it is neither reasonable nor practical to include free-rider savings in counting goals achievement, the proper estimation and accounting for free riders is very important.

Because the Commission will need to include as part of its compliance review an assessment of free-ridership experience and reconciliation of experience with initial planning assumptions, it is critical that utilities provide their program-design assumptions concerning free riders, and that they collect and evaluate program-implementation data on free-rider participation. Otherwise, it will be difficult, if not impossible, to measure program savings or cost-effectiveness.

As discussed below, none of the utilities' DSM plans provide free-ridership assumptions. However, our review of the these plans and of the earlier CEGRR's indicates that actual free-ridership may greatly exceed the levels assumed in the plans. Indeed, actual free-ridership may be needlessly exacerbated by the utilities' sole reliance on RIM-passing strategies. If so, the Commission will need to determine with its retrospective review of program performance whether free-ridership has undermined the utilities' efforts to meet the goals and the extent to which the failure to consider high savings/low rate impact TRC options contributed to the goals shortfall.

A. Goals accounting of free-rider savings

Whether utilities should be allowed to count free-rider savings toward meeting their goals could become a major issue in goals compliance. The Commission's intent on the former seems clear from the case record and the final order in the goals docket: the Commission adopted Staff's recommendation to set goals "based on 100% RIM unadjusted for free riders" (Staff recommendation, September 23, 1994; emphasis added).

Although the intent was unambiguous, the implication for each utilities' treatment of free riders was not. FPC and TECo did not

---

[29]Programs may also prove to be uneconomical in practice if free-ridership is greater than anticipated. Any program expected to have a TRC or RIM cost/benefit ratio of less than 1.13 with an expected free-ridership rate of 10% will in fact not be cost-effective if free-ridership actually amounts to 20%.

account for the effects of free riders in their estimates of potential savings from the RIM portfolio.[30] Instead, they proposed in hearings that for the purposes of setting goals that RIM-portfolio savings be reduced for free riders using a simple adjustment factor. The Commission explicitly rejected this proposal as arbitrary and set goals at 100% of estimated RIM potential. For these utilities, goals were truly unadjusted for free riders.

Gulf and FPL, however explicitly accounted for free-rider savings in their estimates of RIM-portfolio potential.[31] Since the Commission based goals for these utilities on 100% of their "adjusted" RIM-portfolio potential, those utilities' goals implicitly exclude savings from free riders.

The variation across utilities in the treatment of free riders in setting the goals creates ambiguity for the treatment of free-rider savings in meeting the utilities' goals. Moreover, since none of the utilities' DSM plans provide free-ridership assumptions, it is not clear whether utilities believe that free-rider savings can or cannot be counted towards goal achievements. Given this uncertainty, LEAF recommends that the Commission clarify in its orders on the DSM plans that utilities will not be allowed to count free-rider savings to meet goals or to use free-rider savings for purposes of goals compliance.

Allowing utilities to meet goals with free-rider savings would effectively undermine the intent of the Legislature as expressed in the Florida Energy Efficiency and Conservation Act, which directs the Commission to develop and adopt overall goals and authorizes the Commission to require each utility to develop plans and implement programs for <u>increasing</u> energy efficiency and conservation within its service area. Sections 366.80-366.85 and 405.519, F.S.

Utility DSM spending on free-rider participants will not increase energy efficiency in a utility's service territory because free riders are those participants that would have adopted efficiency improvements on their own in the absence of a utility-funded program. Although free riders contribute to improving

---

[30]Nor did they account for free riders in their cost-effectiveness screening of DSM options.

[31]FPL reduced market potential to account for free riders by excluding savings from measures with less than 2-year paybacks and by reducing average savings from residential load management measures (Docket No. 930548-EG, response to People's Gas Interrogatory 1-12). Gulf adjusted for free-ridership by estimating the difference in measure adoption rate in a scenario with no utility DSM spending and one with moderate utility DSM investments (Docket No. 930550-EG, CEGRR (revised 2/22/94), p. 1).

energy efficiency in their utility's service territory, they would do so regardless of the utility's efforts.

Allowing utilities to meet the goals with free-rider savings provides perverse incentives for utilities to game the system. Since free-rider savings will occur without utility intervention, utilities could seek ways to minimize their DSM investments while claiming goals compliance on the basis of free-rider savings.[32]

More critically, counting free-rider savings towards the goals makes the kilowatt-hour a free-rider would have saved anyway as important as the kilowatt-hour that only utility DSM spending can save. Prohibiting utilities from counting free-rider savings appropriately directs utility efforts at maximizing savings that would not otherwise occur--the reason for utility programs in the first place. Excluding free-rider savings thus creates an incentive for utilities to minimize spending on customers that will implement efficiency improvements anyway.[33] Without such an incentive, utilities will spend its customers' money on activities that yield no return in ratepayer benefits.

Including the free riders in the goals would render the goals useless for planning purposes. The utilities typically prepare a base-case load forecast for natural or business-as-usual conditions, including energy efficiency trends caused by changing technology, changing prices, and government regulations. This base-case forecast include the load reductions due to free riders, since by definition free riders are those participants whose load reductions would have occurred without the program. Attempting to remove all free riders (and hence all efficiency improvements) from this base case would be futile and the result would not be a meaningful forecast of anything. LEAF does not believe any utility prepares such a free-riderless forecast.

From this base-case forecast, the utility subtracts DSM effects (interruptible loads, load management, and conservation), along

---

[32]This could be a particularly attractive option if the utility were faced with spending constraints or under pressure to minimize the rate effects of DSM spending.

[33]For example, a utility could operate a program urging commercial customers to use fluorescent, rather than incandescent, lighting in new construction and renovation. Since this measure is decades old and already accepted for 80-90% of commercial floor space, the program would have little effect. The total savings from the measure will be quite large (perhaps high enough to meet the entire DSM goal), but free ridership will be very high, representing close to 100% of savings. If free rider load reductions were counted towards goals compliance, the utility might well meet its goals without actually reducing loads even one MW.

with other adjustments (such as for self-generation), to derive projected net firm loads. If free riders are included in the measurement of DSM, as well as in the base forecast, their effects will be double counted, and the net load requirements will be understated, leading the utilities to under-forecast and under-build. For the goals to be useful in planning, they must represent the reduction in loads from the base case, which requires that they be net of free riders.

If the Commission intends to effect real efficiency improvements and to provide true cost reductions to ratepayers, free-rider savings must be excluded from utilities' claims of goals achievements.[34]

B. Free-Ridership in Utility Savings Estimates

At this time, LEAF is unable to determine the extent to which savings estimates in the utilities' DSM plans have been adjusted for free-ridership. The utilities did not describe free-ridership assumptions in their DSM plans, and they have not yet responded to LEAF's interrogatories regarding free-rider accounting. However, if the utilities estimated free-ridership for the DSM plans in the same manner as in their CEGRRs, it is likely that actual free-ridership will exceed estimated values. In that event, actual non-free-rider savings could fall short of the goals requirement.

For FPC and TECo, actual free-ridership is bound to exceed initial estimates if the DSM plans use the same free-ridership rates assumed in the CEGRRs; the CEGRRs assumed no free riders. Although FPL and Gulf adjusted savings estimates to reflect free-ridership (at least in the CEGRRs), it is likely that these utilities underestimated the magnitude of free-ridership associated with the program delivery strategies relied on in their plans.[35] In particular, these utilities may have drastically reduced measure incentives in order to make measures cost-effective under the RIM.[36]

---

[34]The free-rider savings excluded from claims of goal achievements would include reductions in load due to market-driven switching from electric to gas service. Load reductions from electric-to-gas conversions should be counted for purposes of goal compliance only when such switching results from electric utility spending on efforts to overcome market barriers to cost-effective conversion.

[35]This would also pertain to FPC and TECo if they in fact incorporated free-ridership assumptions in their savings estimates.

[36]Since the RIM only counts utility expenditures as costs, the RIM cost-effectiveness of a measure can be improved by reducing the utility incentive payment and increasing the participant share

Reducing incentive levels, however, increases the share of free-rider participants and savings; since free riders are customers that would install the measure without any incentive, all increases in participation and savings resulting from an increase in incentive level will be from non-free riders. It is unlikely that the utilities reasonably estimated this effect on free-ridership from its reductions in incentive levels.[37]

In essence, the utilities may be needlessly increasing free-ridership by relying solely on RIM-passing measures to meet goals. As a result, the utilities could fall short of their goals simply because they declined the Commission's invitation to consider TRC measures.

To determine whether goals have been met with non-free-rider savings, the Commission's compliance review will need to establish from evaluation results the share of measured program savings attributable to free riders. Based on its review of evaluation studies, the Commission may want to prescribe changes to program strategies to reduce free-ridership, particularly if it finds that free-ridership is undermining utility efforts to comply with the goals requirements and that free-ridership is exacerbated by sole reliance on RIM-passing measures. To ensure that it has sufficient information to assess the effects of free-ridership, the Commission should direct utilities to:

- provide all free-ridership assumptions incorporated in savings estimates in the DSM plans, and provide program-savings estimates both gross and net of assumed free-rider savings;
- report program performance and cost-effectiveness results both gross and net of measured free-rider savings;
- collect all process and impact evaluation data required to verify free-ridership assumptions in the DSM plans.

V. Compliance Review and Enforcement

To hold utilities to their annual goals, the Commission will need to review the performance of the utilities' DSM programs on an annual basis. These annual reviews can serve to enforce the Commission's command and control approach in two ways. First, the

---

of measure cost. The Commission limited this practice by requiring that the measure also be cost-effective under the participant test, which only counts participant expenditures as costs.

[37]However, the utilities have not yet provided the data required to determine whether they accounted for the effect of incentive levels on free-ridership.

compliance reviews would carry the threat of financial penalties for inadequate program achievements in the previous year. Second, the Commission can take timely action to increase the likelihood that future goals are met by prescribing improvements in program strategies.

Neither of these enforcement mechanisms would be as effective if applied less frequently than on an annual basis. Although retrospective penalties could be imposed years after the fact, their effect on utility decision-making would likely diminish as the threat becomes less immediate. Prescriptions for program-design changes could be crafted on the basis of more than one year's worth of implementation experience.[38] However, each year's delay in effecting needed design improvements could mean another year of noncompliance.

Fortunately, the Commission need not open new proceedings solely for the purpose of reviewing goals achievements. Instead, such reviews can be undertaken as part of the annual ECCR proceedings.

There are a number of advantages to coupling an annual prudence review of DSM spending with an assessment of savings achievements from DSM spending. The Commission's considerations of appropriate remedies for inadequate achievements would benefit from evidence on the prudence of a utility's management of its programs; evidence of imprudence could buttress arguments for Commission-mandated program modifications in addition to performance penalties.[39] Alternatively, a review of savings achievements could indicate that DSM budgets were expended on ineffective or non-cost-effective program efforts; such findings could be germane to considerations of management prudence.

Although we agree with the Commission's decision to apply penalties on a case-by-case basis, we recommend that the Commission adopt a consistent approach for determining the magnitude of a performance penalty. Adopting a penalty mechanism would signal clearly the Commission's intent to strenuously enforce its goals directives. A formalized mechanism would also make transparent the

---

[38]In fact, the Commission may want to defer the first round of compliance reviews until results from impact evaluations are available. Full impact evaluations may take eighteen months to two years to complete. However, interim results may be available in a shorter time frame, especially if specialized evaluation techniques are employed. See the discussion of implementation analyses above.

[39]However, the decision to impose a noncompliance penalty should not be contingent on a finding of imprudence. Conversely, cost disallowances could be warranted in conjunction with performance penalties if program costs were imprudently incurred.

utilities' exposure to financial penalties if they fail to comply with the goals mandates.[40]

Finally, a properly designed mechanism could be applied symmetrically to provide either penalties for noncompliance or shareholder incentives for exceeding goals with TRC measures or programs that offer large savings with small rate effects. For example, shareholder incentive mechanisms have been adopted in several states that base the incentive on a share of the net economic benefits that result when pre-determined goals are exceeded. In some states, these mechanisms are designed to impose symmetrical penalties equivalent to a share of the net economic loss resulting from a failure to realize pre-set goals.

These shared-savings mechanisms generally rely on straightforward methods to calculate and apply an incentive or penalty for exceeding or falling short of a pre-set goal. Gross benefits are measured as the difference in avoided supply (generation, T&D, reserve) costs avoided by the savings in excess of the goal.[41] Net benefits are simply gross benefits less any additional DSM spending to acquire the excess savings. The shareholder incentive is then set as a percentage of total net benefits.[42] Where savings fall short of goals, the penalty calculation mirrors that for an incentive.

The exact form of the penalty-incentive mechanism for goals compliance does not need to be finalized at this time. Instead, the Commission should formally designate the annual ECCR proceedings as the venue for its compliance reviews and indicate its intent to consider proposals for penalty-incentive mechanisms during the first round of reviews. A directive from the Commission in this regard will serve notice on the utilities of the Commission's commitment to finding equitable means for penalizing inadequate achievements and rewarding exemplary performance.

---

[40]Adopting a formal penalty mechanism would not limit the Commission's authority to impose additional penalties or disallowances for management imprudence.

[41]The gross benefits are typically calculated as the present value of avoided costs over the life of the excess savings.

[42]In some jurisdictions, the percentage share of net benefits or loss increases with the amount that savings exceed or fall short of the goal.

ORNL/CON-336

HANDBOOK OF EVALUATION OF UTILITY DSM PROGRAMS

ERIC HIRST AND JOHN REED
Editors

BENSON BRONFMAN, GEORGE FITZPATRICK, ELIZABETH HICKS, ERIC HIRST, MARC HOFFMAN, KENNETH KEATING, HARVEY MICHAELS, STEVEN NADEL, JANE PETERS, JOHN REED, WILLIAM SAXONIS, ANDREW SCHÖN, AND DANIEL VIOLETTE
Authors

DECEMBER 1991

Sponsored by
New York State Energy
Research and Development Authority
David R. Wolcott, Project Manager
and
Office of Conservation and Renewable Energy
U.S. Department of Energy
Diane Pirkey, Project Manager

OAK RIDGE NATIONAL LABORATORY
Oak Ridge, Tennessee 37831
managed by
MARTIN MARIETTA ENERGY SYSTEMS, INC.
under contract No. DE-AC05-84-OR-21400
for
U.S. DEPARTMENT OF ENERGY

PREFACE

Diane Pirkey
U.S. Department of Energy

and

David R. Wolcott
New York State Energy Research and Development Authority

We are pleased to have the opportunity to produce this handbook on the evaluation of utility demand-side management (DSM) programs. This volume is the result of a research project that was cosponsored by the Integrated Resource Planning programs of the New York State Energy Research and Development Authority and the U.S. Department of Energy.

Program evaluation has become a central issue in the world of utility integrated resource planning. The DSM programs that utilities were operating to meet federal requirements or to improve customer relations are now becoming big business. DSM is being considered an important resource in a utility's portfolio of options. In the last five years, the amount of money that utilities have invested in DSM has grown exponentially in most regulatory jurisdictions. Market analysts are now talking about DSM being a $30 billion industry by the end of the decade. That is enough money to warrant serious efforts to determine how effectively utilities are carrying out the mission of implementing energy efficiency in their customers' facilities.

If the large volume of DSM-program investments was not enough to highlight the importance of evaluation, then the introduction of regulatory incentives has really focused the spotlight. Many regulators now understand the importance of aligning a utility's profitability with integrated resource planning goals as the means of obtaining the enthusiastic support of utility management for DSM. Various mechanisms have evolved to allow a utility to recover its costs and earn a bonus for successfully and aggressively acquiring DSM resources. The most popular mechanisms pay the utility either a "bounty" per unit of energy saved or a "shared savings" percentage of the total resource savings generated by a DSM program. In both cases, the incentive is based on the amount of energy saved, which can only be determined by some evaluation method. In this context, evaluation has gained the attention of all the participants in utility rate cases, from regulatory accountants to ratepayer advocates. Evaluation has come of age!

This handbook was developed through a process that involved many of those people who represent the diverse constituencies of DSM-program evaluation. For example, a team was

ORNL/CON-336

HANDBOOK OF EVALUATION OF UTILITY DSM PROGRAMS

ERIC HIRST AND JOHN REED
Editors

BENSON BRONFMAN, GEORGE FITZPATRICK, ELIZABETH HICKS, ERIC HIRST, MARC HOFFMAN, KENNETH KEATING, HARVEY MICHAELS, STEVEN NADEL, JANE PETERS, JOHN REED, WILLIAM SAXONIS, ANDREW SCHÖN, AND DANIEL VIOLETTE
Authors

DECEMBER 1991

Sponsored by
New York State Energy
Research and Development Authority
David R. Wolcott, Project Manager
and
Office of Conservation and Renewable Energy
U.S. Department of Energy
Diane Pirkey, Project Manager

OAK RIDGE NATIONAL LABORATORY
Oak Ridge, Tennessee 37831
managed by
MARTIN MARIETTA ENERGY SYSTEMS, INC.
under contract No. DE-AC05-84-OR-21400
for
U.S. DEPARTMENT OF ENERGY

TABLE OF CONTENTS

PREFACE

Diane Pirkey
U.S. Department of Energy

and

David R. Wolcott
New York State Energy Research and Development Authority

We are pleased to have the opportunity to produce this handbook on the evaluation of utility demand-side management (DSM) programs. This volume is the result of a research project that was cosponsored by the Integrated Resource Planning programs of the New York State Energy Research and Development Authority and the U.S. Department of Energy.

Program evaluation has become a central issue in the world of utility integrated resource planning. The DSM programs that utilities were operating to meet federal requirements or to improve customer relations are now becoming big business. DSM is being considered an important resource in a utility's portfolio of options. In the last five years, the amount of money that utilities have invested in DSM has grown exponentially in most regulatory jurisdictions. Market analysts are now talking about DSM being a $30 billion industry by the end of the decade. That is enough money to warrant serious efforts to determine how effectively utilities are carrying out the mission of implementing energy efficiency in their customers' facilities.

If the large volume of DSM-program investments was not enough to highlight the importance of evaluation, then the introduction of regulatory incentives has really focused the spotlight. Many regulators now understand the importance of aligning a utility's profitability with integrated resource planning goals as the means of obtaining the enthusiastic support of utility management for DSM. Various mechanisms have evolved to allow a utility to recover its costs and earn a bonus for successfully and aggressively acquiring DSM resources. The most popular mechanisms pay the utility either a "bounty" per unit of energy saved or a "shared savings" percentage of the total resource savings generated by a DSM program. In both cases, the incentive is based on the amount of energy saved, which can only be determined by some evaluation method. In this context, evaluation has gained the attention of all the participants in utility rate cases, from regulatory accountants to ratepayer advocates. Evaluation has come of age!

This handbook was developed through a process that involved many of those people who represent the diverse constituencies of DSM-program evaluation. For example, a team was organized in New York, composed of regulatory staff, environmental advocates, utility personnel, and state energy policy planners. The group selected the handbook topics, suggested chapter authors, and produced a successful conference that brought the chapter authors together with almost 200 DSM practitioners. The resulting "chemistry," involving many different points of view, had a lot to do with the success of this handbook.

We have come to recognize the many technical disciplines that must be employed to evaluate DSM programs. An analysis might start out based on the principles of utility load research to find out what happened, but a combination of engineering and statistical methods must be used to "triangulate" an estimate of what would have happened without the program. The difference, of course, is that elusive but prized result of evaluation: what happened as the direct result of the DSM program. Technical performance of DSM measures is not the sole determinant of the answer, either. We also recognize the importance of such behavioral attributes of DSM as persistence and free ridership. Finally, DSM evaluation is meaningless without attention to planning an approach, communicating results to relevant decision-makers, and focusing as much on the process as the impacts of the program. These topics are all covered in this handbook.

We were honored that each of the authors of chapters in this handbook was our number one pick as the most authoritative person on each of these particular topics. This is a world-class group of authors, made more so by our editors, Eric Hirst and John Reed. We are indebted to both of them for having the vision to see the need for this volume, the technical competence to "pull off" a credible result, and the perseverance and discipline to produce such a professional product. We are sure you will agree that this handbook makes a significant and original contribution to the field of DSM-program evaluation.

LIST OF ACRONYMS

ACEEE	American Council for an Energy Efficient Economy
BPA	Bonneville Power Administration
CI	Confidence interval
C&I	Commercial and industrial
CMP	Central Maine Power Company
DOE	U.S. Department of Energy
DSM	Demand-side management
EDA	Exploratory data analysis
EIA	Energy Information Administration
EPRI	Electric Power Research Institute
ESCO	Energy service company
HSEM	Hybrid statistical/engineering model
HVAC	Heating, ventilating, and air conditioning
NEES	New England Electric System
PC	Personal computer
PG&E	Pacific Gas & Electric Company
PRISM	Princeton Scorekeeping Method
PSC	Public service commission
PUC	Public utility commission
R&D	Research and development
RFP	Request for proposals
SAE	Statistically-adjusted engineering
SD	Standard deviation
SH	Space heating
SIC	Standard Industrial Classification
WAP	Weatherization Assistance Program
WH	Water heating

INTRODUCTION

Eric Hirst and John Reed

BACKGROUND

The 1990 annual report from Consolidated Edison (1991), which serves three million customers in New York, states that the Company's goal

> ... is to provide reliable electric service at a competitive cost, in an environmentally compatible manner, while providing a fair return for our stockholders. Energy efficiency and diversity are keys to achieving this goal. During 1990 we dramatically expanded our programs to help customers use energy more efficiently and thereby limit growth in electricity usage. ... The company expects to spend $140 million on these programs over the next two years. ... Our aim is to reduce usage of electricity by about 15 percent from what it otherwise would be by the year 2008.

Utilities and state public utility commissions (PUCs) in other parts of the country are making similar commitments to using DSM programs. Utilities in Maine, Massachusetts, Connecticut, New York, Wisconsin, Washington, Oregon, and California use DSM programs as alternatives to building new power plants. For example, the New York utilities will spend $4 billion on DSM programs during the 1990s (Bradford 1991). Niagara Mohawk, Consolidated Edison, and Long Island Lighting, the three largest electric utilities in New York, plan to reduce projected electricity consumption by the year 2000 by 5%, 8%, and 9%, respectively [New York Public Service Commission (PSC) 1990]. And these utilities plan to cut summer peak demands by 10%, 13%, and 11%, respectively, that year.

As these DSM programs grow in size, cost, and expectations, measuring their effects and cost-effectiveness is increasingly important. Providing credible answers to certain key questions is vital: Are DSM programs providing the energy savings and load reductions expected of them? Are these programs cost-effective for utility customers? What can be done to increase program participation, lower costs, and boost energy and demand reductions?

Evaluation, the subject of this handbook, is the process that provides answers to these and other questions about the operation and effectiveness of DSM programs. The purpose of this handbook is to provide semitechnical guidance on how to plan and conduct evaluations and to point out the pitfalls to avoid. It is written for midlevel utility staff who manage and evaluate DSM programs. It also should be valuable to users of evaluation results. In particular, staff at PUCs, environmental groups, and other organizations participate in the planning and review of utility DSM programs; this handbook should help them become well informed reviewers and consumers of evaluation plans and evaluation results.

WHAT IS EVALUATION?

Evaluation is the systematic measurement of the operation and performance of programs (Hirst 1990a). Evaluations depend on objective measurements rather than anecdotal evidence and personal impressions. Evaluations use social-science research methods and technical data to produce valid and reliable results. And evaluations are intended to affect future decisions; they are not academic exercises (Kushler 1989).

Evaluations provide information for program managers and staff that they can use to modify and improve program operations. More broadly, evaluations provide information needed by utility executives and regulatory agencies to assess these programs. The results of evaluations

- Document the energy savings, load reductions, and cost-effectiveness of DSM programs
- Show how to improve programs by increasing participation rates, raising energy savings, or cutting costs
- Suggest ways to improve the design of future programs
- Support DSM budgets before the utility's budget committee
- Provide data to improve the company's load forecasts and resource planning

Utilities should measure the performance of DSM programs with the same competence and diligence with which they monitor the performance of power plants. During the past several decades, utilities have developed detailed information on construction costs and time; operation, maintenance, and fuel costs; heat rate, availability factor, and capacity factor; the duration and causes of each outage; and fuel consumption, plant output, and emissions for each of their power plants. Because the history of DSM programs is so short, comparable data generally do not exist for energy-efficiency and load-management programs on program-participation rates, energy savings (GWh), load reductions (MW), and program costs (Hirst and Sabo 1991).

TYPES OF EVALUATION

Evaluators typically speak of two types of evaluation. *Process* evaluations examine program operations to identify how well the program is implemented and to suggest ways to improve program delivery (Fig. 1). Such evaluations focus on program goals, history, and activities and often are based on interviews with utility program staff, program managers, participants, and trade allies.

Because process evaluations usually occur during program implementation, they can provide prompt feedback to help improve program design and implementation. In addition, process evaluations provide important insights that can be used to interpret findings from impact evaluations.

Impact evaluations examine the effects of the program (Fig. 1). They provide quantitative documentation of program benefits and costs. Impact evaluations measure program participation, participant acceptance of the recommended DSM measures and practices, performance of the DSM technologies promoted by the program, program energy and load reductions, and program costs.

Impact evaluations compare what happened to program participants with what would have happened to participants if the program had not existed. These evaluations deal with two aspects of energy savings and load reductions. *Total* savings are the changes in annual electricity use and peak demand experienced by participants in the utility's program. *Net* savings are the portion of the total savings that can be directly attributed to the utility program (Fig. 2).

Net savings are the difference between total savings and the savings that participants would have achieved had the program not existed. Nonprogram savings reflect customer responses to changes in electricity and fossil-fuel prices, changes in economic activity or personal income, introduction of new electricity-using technologies, and other nonprogram factors. While it is possible to measure directly total savings, nonprogram savings cannot be measured because the program exists. Estimation of nonprogram savings requires a comparison group to approximate what program participants would have done without the program. Nonparticipating customers that are eligible for the program are often used as a comparison; sometimes, preparticipation data for participants is used as a comparison. Using data from nonparticipants and preprogram data from participants can often give the most reliable estimates of net effects.

Energy savings and load reductions are determined primarily by analyzing monthly electricity bills and load-research data. These electricity-use data are often supplemented with data on weather, occupant and equipment characteristics, facility characteristics, and program records. Depending on the purpose of the programs, evaluations focus on energy savings (GWh), peak-load reductions (MW), or changes in load shape.

THE EVALUATION PROCESS

Table 1 lists the activities associated with an evaluation. First, the purposes of the DSM program and the objectives of the evaluation need to be clearly identified. The goals of the evaluation depend on the goals of the program. In particular, the evaluation should focus on topics that will inform future decisions about the program. Evaluation goals, in turn, affect the design of the evaluation. If, for example, the program is just beginning, the evaluation will probably focus on program process because an impact evaluation would be premature. If, on the other hand, a pilot program is ready to be implemented systemwide and the expanded program is expected to contribute substantial energy or capacity resources, then an impact evaluation is appropriate.

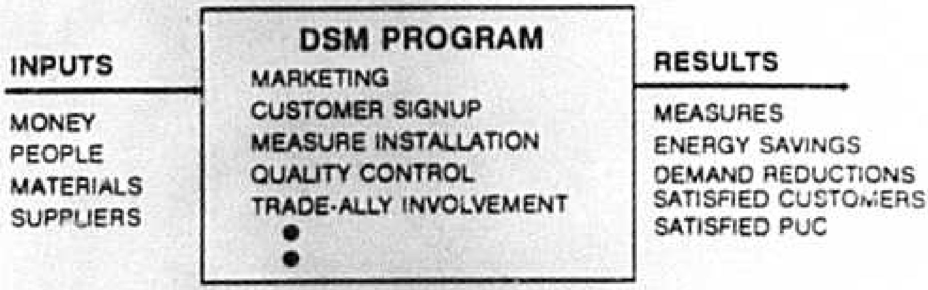


Fig. 1. Relationships among a DSM program, its inputs, and outputs. Process evaluation deals with the activities inside the box. Impact evaluation deals with the relationships between program inputs and results.

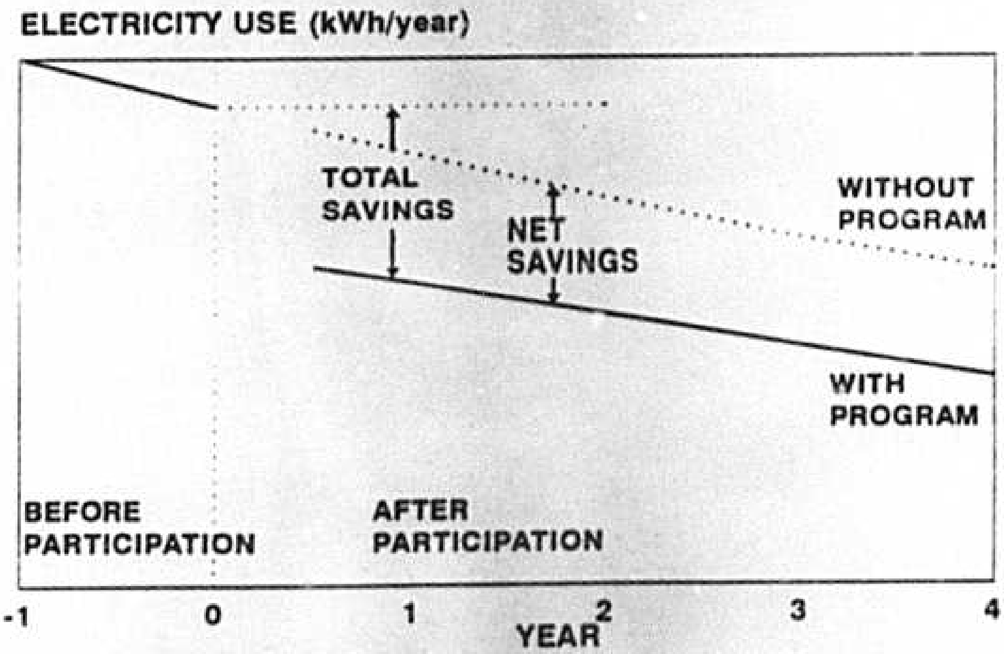


Fig. 2. Schematic of electricity use per participant for commercial buildings with and without a utility program aimed at improving heating, ventilating, and air conditioning (HVAC) systems. In this example, preparticipation electricity use is declining in response to increasing electricity prices.

Table 1. **Sequence of activities in evaluating DSM programs**

Plan evaluation
Identify program goals and future decisions
Determine purposes of evaluation and questions to address
Identify evaluation resources (budget, staff, time, data, and analytical methods)
Implement evaluation
Define populations to sample
Collect data (program records, weather data, monthly electricity bills, customer surveys, interviews with program staff, and end-use load monitoring)
Analyze data
Present evaluation results
Review results with program implementation and planning staffs
Report results to utility executives and to the PUC
Prepare final report and briefings
Act on findings

The level of effort (dollars and staff) devoted to the evaluation needs to be determined at this stage. The evaluation should reflect the importance of the program itself. Evaluations need to be carefully planned so they address the major issues of the particular DSM program. Evaluations also should examine alternative explanations for what occurred (called threats to validity in the evaluation literature). Such threats include history (another event, such as an increase in electricity prices, that was responsible for the observed changes), selectivity bias (program participants and nonparticipants may differ in ways that affect their electricity use), and instrumentation (errors in the methods used to measure changes in electricity use). Finally, the evaluation plan should identify the data that will be collected, by whom, how, and when.

Implementing an evaluation includes the selection of samples of customers, utility employees, and vendors to interview and the development and testing of the questions to ask them. Other data sources, such as utility billing records, local weather data, and program records are identified and collected at this stage. The objectives of the program and its evaluation will suggest what types of data are most important. If the program is aimed at reducing demand during system peak (e.g., a direct load-control program), then load-research data may be required. If the program is intended to improve overall energy efficiency (e.g., through rebates for efficient appliances and lighting), then monthly billing data are suitable. These data then need to be merged, checked for errors, and analyzed.

Finally, evaluation results are prepared. This involves first a review of the results with program staff and management. Then results are presented to utility executives, and reports are prepared for internal and external distribution.

EVALUATION SKILLS

The evaluation group in a utility should include professionals with diverse educational backgrounds and experience. Appropriate skills include market research, economics, psychology, sociology, statistics, engineering, and business. The group should possess a mix of personal, communication, technical, and analytical skills to deal with the diversity of issues that arise in planning and conducting evaluations.

Technical skills are required to examine and use alternative data sources, including energy audits that identify cost-effective DSM options, and special meters and wiring used to measure end-use load-shapes. Analytical skills are needed to address a variety of potential problems that can confound interpretation of evaluation data. These problems include accuracy of self-reports in surveys, nonresponse to surveys, self-selection into the program,[1] and differences between early and later participants. Statistical skills are needed to define sample frames and suitable sampling methods and to define appropriate sample sizes. Computing skills are needed to manage and ensure the quality of the large and diverse data sets created for evaluations. Personal and communication skills are needed because collection of evaluation data and application of evaluation results depend on the cooperation of people in other departments within the utility.

A recent study of energy savings stimulated by retrofits in commercial buildings illustrates the complications that can occur in interpreting evaluation data and results and, therefore, the importance of capable evaluators (Exhibit 1).

EMERGING TRENDS

Surveys conducted by the Electric Power Research Institute (EPRI) show that U.S. electric utilities are conducting more than 1300 DSM programs. In 1990, these programs cut annual electricity use by 1.3% and cut summer peak demand by 3.7% (Faruqui et al. 1990). The energy and load impacts of utility programs are likely to increase dramatically during the next two decades (Hirst 1991). Several factors that affected utility DSM programs in the 1980s will change dramatically during the 1990s to increase the size, scope, cost, and effectiveness of DSM programs, including:

- Growing interest in integrated resource planning, which involves explicit consideration of DSM programs as cost-effective alternatives to some new power plants

- Increasing public concern about the environmental effects of electricity production and transmission, especially global warming and acid rain

---

[1]Because participation in most programs is voluntary, there may be important differences between those who choose to participate and those who choose not to participate.

Exhibit 1. Engineering audits, billing data, and end-use data yield different interpretations

Puget Energy Services (1991) used engineering analyses of the measures actually installed, monthly billing data, and end-use load data to estimate electricity savings for a few commercial buildings in the Pacific Northwest. These buildings had been retrofitted as part of a conservation program run by the Bonneville Power Administration (BPA). Although, on average, the three sets of estimates agreed closely, the differences for individual buildings were quite large (Fig. 3). For only five of the eleven buildings with end-use load-research data did the engineering predictions fall within 20% of the "actual" savings. Three of the buildings showed *negative* savings, based on analysis of monthly electricity bills.

These differences in estimates of energy savings are less a question of finding the correct answer than of defining the appropriate question to ask. For example, monthly billing data and end-use data measure different properties. Billing data measure changes in electricity use for the building as a whole, whereas the load-research data measure changes in electricity use for particular end uses. Thus, changes in equipment and operating practices for end uses not covered by this retrofit program show up in the billing data but not in the load-research data or the engineering analyses.

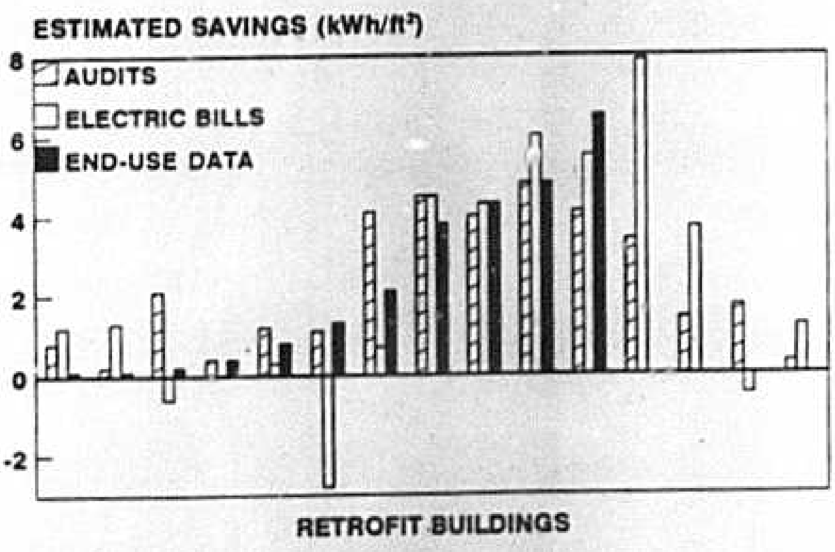


Fig. 3. Estimates of electricity savings in 15 retrofit commercial buildings in the Pacific Northwest (Puget Energy Services 1991).

- Provision of financial incentives to utility shareholders for implementing cost-effective DSM programs

- Growing recognition of the powerful role that utilities can play in overcoming market barriers to the adoption of cost-effective energy-efficiency opportunities

Perhaps the most important of these changing trends (and the one most likely to affect evaluations dramatically) is the provision of financial incentives. Such regulatory mechanisms already exist in Massachusetts, Rhode Island, New Hampshire, New York, and California and will soon be adopted in several other states. All seven utilities in New York State have incentive systems in place (Gallagher 1991).

The most popular incentives use shared-savings mechanisms, in which the utility keeps part of the net benefit provided by its DSM programs (Fig. 4). The net benefit is the difference between the total benefits and program costs. Total benefits are typically defined as the amount of energy saved by the program multiplied by the avoided energy cost plus the amount of demand reduction multiplied by the avoided capacity cost. The critical element in computing net benefits is estimating energy and demand reductions, the province of program evaluators.

Consider a hypothetical commercial lighting program as an example of the ambiguities in a carefully conducted evaluation. The program is aimed at medium-size office buildings and includes general information and onsite lighting audits. These activities identify suitable lighting measures and encourage the customer to apply for the 50% rebate offered by the utility. The rebate helps defray the costs of energy-efficient lamps, ballasts, fixtures, and controls.

The utility ran a comprehensive evaluation that included three elements, using approaches discussed by Violette et al. (1991) and XENERGY (1990):

- Analysis of two years of electricity billing data, one year before participation and one year after, for samples of participants and eligible nonparticipants

- Thirty days of time-of-use metering, pre- and postretrofit, of a sample of lighting circuits for a sample of participants only (no comparison group)

- Engineering analysis of the energy and load reductions caused by the measures actually installed by participants (again, no comparison group)

The utility used multiple methods to estimate program savings because each method is imperfect. If the utility knows how the results of these disparate methods will be used to determine program effects, such triangulation can build confidence in the estimates ultimately used. Not surprisingly, these approaches and their associated analytical procedures gave different estimates of net energy savings (Table 2); for simplicity, the load-reduction effects of the program are ignored in this example.

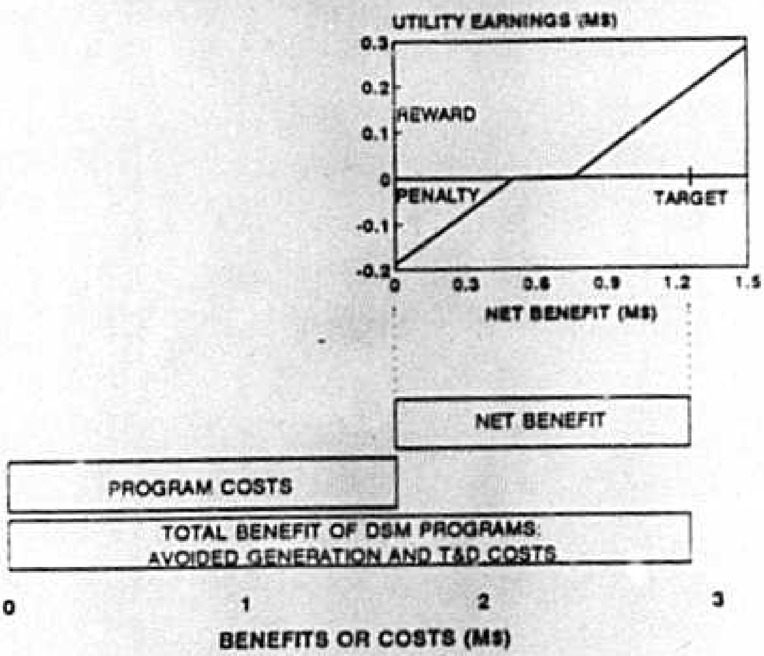


Fig. 4. **Schematic showing the mechanics of a shared-savings mechanism to reward utility shareholders for implementing cost-effective DSM programs. In this example, the program is expected to cost $1.6 million, the total benefits are expected to be $2.9 million, and the net benefit is then $1.3 million. Shareholders earn $0.19 million if the utility achieves its net-benefit target of $1.3 million.**

Comparison of pre- and postretrofit electricity use, based on monthly billing data, for participants showed a reduction of 9600 kWh/year per rebate participant. Nonparticipants showed a small increase in electricity use during the same two-year period. (The local economy was growing during this period, which led to higher occupancy levels and longer hours of operation for office buildings. These changes in building use explain the increase in electricity use for nonparticipants.) And the customers who received a lighting audit but did not apply for the rebate (one-third of the total number of customers that received audits) cut their consumption slightly. Table 2 shows the roughly 50% difference in estimated annual electricity savings, based on analysis of billing data. These differences depend on whether participant savings are adjusted for nonparticipant changes in electricity use and for the savings achieved by audit-only participants.

The data from ad hoc metering showed savings about the same as those from analysis of the billing data for the participants only. Complications arose in scaling up the metering results to a full year. The metering covered roughly 60 days during a three-month period, with the middle month devoted to installation of new lighting measures. The amount of electricity used for lighting varies with season, and is larger in the winter than in the summer. On the

the net benefit exceeds $753 per rebate participant. Again, following Fig. 4, the incentive fraction is 38% of the net benefit above the 60% threshold level. Thus, the structure of the incentive mechanism adds more leveraging in going from net benefits to utility incentive.

While the use of evaluation results in contested regulatory proceedings may become a particularly difficult issue for evaluators, many other challenges are likely to face utility evaluators (Table 3). For example, as the budgets and staffing for evaluations increase, it will become more important to coordinate evaluations with other activities in the utility, including market research, load research, load forecasting, DSM-program planning, and integrated resource planning.

Table 3. **Challenges likely to face evaluators of utility DSM programs in the 1990s**

Role of evaluation in contested PUC hearings
Integration of evaluation with other customer-analysis activities
Measurement of the long-term durability of energy savings
Definition of key DSM-program terms
Development of standard approaches for evaluation
Evaluation of the performance of DSM technologies
Measurement of the savings from DSM-bidding programs
Evaluation of DSM programs that aim to transform the market

HANDBOOK CONTENTS

Chapters 2 through 5 are organized around the evaluation process shown in Table 1. Elizabeth Hicks describes the steps associated with evaluation planning in Chapter 2. Harvey Michaels, Marc Hoffman, and Andrew Schön explain the various types of data that are collected for evaluations in Chapter 3. Daniel Violette presents alternative engineering, statistical, and combined techniques to analyze evaluation data in Chapter 4. And Steven Nadel explains how the voluminous evaluation data and analyses can be presented to different groups in Chapter 5.

Chapters 6 through 9 discuss several critical topics that pervade evaluations of DSM programs. Kenneth Keating presents data and suggests methods to measure the long-term energy and load reductions caused by utility DSM programs in Chapter 6. George Fitzpatrick discusses end-use load-research data (data that are particularly detailed, valuable,

PLANNING EVALUATIONS

Elizabeth Hicks

INTRODUCTION

This chapter discusses the critical steps in planning an evaluation (Exhibit 2). Perhaps the key is identifying evaluation goals. Of necessity, evaluation goals must be closely linked to program goals. The various stakeholders in an evaluation may have different goals and may differently perceive the importance of shared goals. Obtaining agreement on evaluation goals among the stakeholders is one way of building commitment to evaluation and the use of evaluation results.

Marshaling resources is an important part of any evaluation plan. One aspect is identifying baseline information requirements and locating the available information. Another part of the resource puzzle is mobilizing information-system resources. Determining budget and staffing is a significant element in any plan. The skills of the available staff will determine whether evaluations are done in-house or are contracted. Successful evaluations usually require the coordination of resources among many departments within the utility, which in turn requires a good deal of planning and political skill.

Selecting an evaluation approach is an important step. Which evaluation approach is chosen depends on a variety of factors, including the size of the program, the life cycle stage of the program, the available resources, the goals of the program, the goals of the evaluation, how the evaluation will be used, regulatory interest, the possibility of collaborative efforts, and the characteristics of the utility (including future capacity needs and the extent to which evaluation results may influence rates and incentives). Finally, developing a research design is a critical part of the evaluation-planning function. The research design identifies when measurements are made, which measurements are taken, and how the measurements are analyzed. A good research design will help to unambiguously answer the question of whether a program has been effective by ruling out alternative explanations for any observed effects.

EVALUATION GOALS

Evaluation needs should be considered at the initial stages of program design. Program design will influence evaluation requirements, the type of evaluation conducted, and the extent of evaluation activities. And evaluation design may influence the nature and timing of program activity (e.g., the type of data collected and the timing of data collection).

Exhibit 3. Relating evaluation goals to particular tasks

As part of a three-year, multimillion dollar evaluation of the U.S. Department of Energy (DOE) Weatherization Assistance Program (WAP), Beschen and Brown (1991) prepared a detailed evaluation plan. This plan began with enumeration of the purposes for this major evaluation. These evaluation objectives led, in turn, to establishment of several studies to address one or more of the objectives (Table 4). For example, the impact evaluation of single-family homes weatherized in 1989 will examine energy and nonenergy impacts, cost-effectiveness, factors that account for program success, and the remaining potential for energy savings.

Table 4. Relationship of evaluation goals to five studies in evaluation of the DOE low-income Weatherization Assistance Program

	Principal submarket studies			Related studies	
	Fuel oil study	Single-family study	High-density multifamily study	WAP network characterization	Eligible client profile
Estimate energy savings	●	●	●		
Assess nonenergy impacts	●	●	●		
Assess cost-effectiveness	●	●	●		
Analyze contributing factors	●	●	●	○	○
Characterize eligible population	○	○	○		●
Describe WAP network capabilities	○	○	○	●	○
Define promising opportunities	○	●	●	●	●

○ = Minor focus ● = Major focus

in maximizing revenues or making the company appear responsive to environmental concerns, a vice president may be interested in increasing control over program resources, and a program manager may be interested in demonstrating effectiveness as a program manager. All parties with a stake in the evaluation should be involved in setting the goals of the evaluation. This involvement may be accomplished by joint goal setting, collaborative processes, or regulatory proceedings. Changes in goals or in the scope of proposed evaluations need to be communicated in a timely fashion by the evaluators to the other parties (Hicks 1990).

Evaluations, particularly process evaluations, are inherently political. A clear understanding of the issues to be examined in a particular evaluation will likely reduce potential conflicts and will contribute to constructive comments as well as to the use of the evaluation results (McRae 1990). If evaluation results are to be accepted and used, all interested parties must be involved in setting the evaluation goals.

At the utility level, this policy implies that evaluation, planning, and implementation staff work jointly to establish evaluation goals, requiring a mutual understanding of program and evaluation tasks and processes. Each evaluation must be tailored to the individual program (Rossi and Freeman 1982). Outside parties who have been involved in program design as part of a collaborative planning processes and/or parties who may review programs and program results, such as regulators, should also be involved in developing evaluation goals. If evaluation results will be used to determine a financial incentive for the utility, all interested parties should jointly develop guidelines for this procedure and agree on the samples, data, and analytic procedures that will be used to determine the incentive.

How the evaluation results are to be used should be factored into the planning of the evaluation and the setting of goals (Chapter 9). Evaluations can provide feedback on program performance, improve planning estimates, and provide cost-benefit analyses of DSM programs. An emerging trend is to use impact-evaluation results to establish utility DSM-incentive payments. Implementers often have specific questions (e.g., what will be the effects of different rebate levels on participation rates or cost-effectiveness?) that need to be examined in the evaluation.

Utilities operate different types of programs, research and development (R&D) programs, pilot programs, full-scale innovative programs, and mature system-wide programs. Evaluation goals will vary with the type of program.

In an R&D program, a small number of customers is treated, and the results are measured carefully. Generally, the objective of this type of program is to develop better estimates of impacts of a technology or a delivery mechanism and to identify potential implementation problems. An example of such a program might be a space-heating (SH)-control experiment conducted in 50 homes with end-use metering. The purpose of the experiment might be to see how SH could be controlled without causing significant discomfort to customers. This type of experiment can give a good indication of the potential gross demand savings, but it would not be useful for determining free-ridership levels. Detailed monitoring of specific technologies, testing of delivery mechanisms, and innovative rate experiments are other types of R&D programs.

A pilot program is a moderately-scaled version of a program with several hundred to several thousand participants, which later may be run at full scale. The purpose of a pilot program is to improve the organizational efficiency of a program and to identify organizational and technological problems that could prove costly if the program is run at full scale. Process-evaluation results may be used to fine-tune the full-scale program design. Preliminary impact evaluations are performed with traditional methods, such as billing analysis, although the program may become a full-scale program before an impact evaluation is completed.

A full-scale innovative program is one that is open to all customers of a given class and that the utility is still refining. The refinements may include adding or deleting measures, changing the incentives, or providing more services to the customer. Either process- or impact-evaluation techniques might be used to evaluate this type of program. And these programs present an interesting challenge because the results from an evaluation can quickly become dated if the program changes. To illustrate, if the incentive structure for a program is changed, earlier free-rider estimates will no longer be valid.

Steady-state programs are those that have been operating for several years and have reached the point that program changes are minor and several impact evaluations have been completed with consistent results. Evaluations for this type of program may be focused on fine-tuning the program's operation and/or on examining the persistence of savings. Evaluations for this type of program may not require large amounts of resources and may largely be based on records maintained by program implementers and the billing department. Very few utility programs currently are at this stage although a few examples exist, such as Seattle City Light's multi-family retrofit program (Okumo 1990).

Many evaluation studies, regardless of cost or sophistication, are not used by practitioners (Franklin and Thrasher 1976). Part of the reason may be the lack of management commitment to evaluation. Obtaining management commitment is usually easier if the evaluator can provide evaluation results in a timely fashion and if the results are relevant to management decision making. In such a situation, the evaluator must anticipate decision-maker information requirements and the time frames in which decisions must be made. The evaluator must communicate with decision makers about their needs and may need to help them analyze their information requirements. The timeliness issue may shape the nature of the evaluation because it may require the use of results from other studies, it may limit efforts to a process evaluation, or it may limit impact measurements to a small sample. If management is committed to evaluation, the staff, time, and money needed to perform good evaluations will be made available, and evaluation findings will be used.

A major consideration in setting evaluation goals is determining the characteristics of the utility system. The evaluator needs to keep in mind such things as whether a utility is summer or winter peaking, whether it is energy or capacity constrained, and whether avoided costs vary significantly with time. These factors will guide how much emphasis to place on measuring energy or demand impacts and whether time-differentiated impacts are needed.

For example, two utilities may offer a similar commercial lighting program. Utility A is capacity constrained and has very large time-of-day and seasonal differentials in its avoided costs. Utility B does not need capacity for 15 years; most of its baseload generation is coal and it has a very small differential in avoided costs between on and off peak. Utility A may focus its evaluation on when the savings occur during the day and season while utility B may only need to determine total kWh impacts with little regard for load-shape effects.

BASELINE INFORMATION

A crucial element in planning for evaluations is taking stock of the internal and external sources of information about a utility's customers. Basic customer information is particularly useful in planning and evaluating DSM programs. At the very least it is useful to have load data by customer class or by building type as well as information on appliance characteristics and saturations and commercial and industrial (C&I) building stock (Fig. 5). This information, if not kept in the departments that perform DSM planning or evaluation, might be found in the load-research, market-research, or load-forecasting departments. Other possible sources include neighboring utilities, EPRI, the Energy Information Administration (EIA), the Bureau of the Census, and state agencies, especially state energy offices. If this information is not available, the evaluation group may have to collect it, and the cost of this data collection will need to be factored into budget planning.

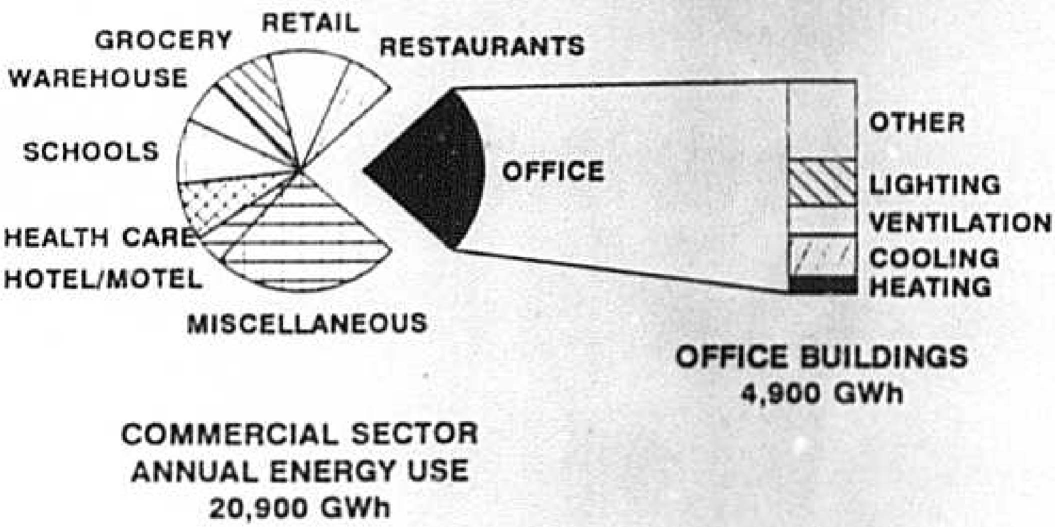


Fig. 5. Data and estimates on annual electricity use by building type and end use for the commercial sector in the Georgia Power service area. This information is important for DSM technology and program screening because it helps to identify the market segments with the largest potentials.

PLANNING FOR INFORMATION-SYSTEM REQUIREMENTS

Information systems that contain data on program activities and installations associated with a particular program are important to any large-scale evaluation effort. Depending on the utility and the nature of the program (R&D, pilot, or full-scale), information-tracking systems may be program specific or part of the larger corporate information-management systems.

The program evaluator needs to be very careful about creating a program-specific standalone information system that is independent of other corporate systems because evaluation data often need to be integrated with other corporate databases.

Generally, data from program databases and the corporate information system are needed to make initial estimates of program savings. Ideally, these systems contain information on such items as rate class, building type, Standard Industrial Classification (SIC) code, measures installed, hours of facility operation, and costs of measures, information that is essential to process and impact evaluations.

Evaluation requirements should be considered when program databases are being developed. Often this foresight is not practiced, and evaluation information is collected later through different systems. This difference can lead to inconsistencies that are difficult to reconcile during the evaluation. Evaluators should work closely with program implementers and information-systems people to ensure that their data-collection needs are met. Evaluators should also perform validity and quality-control checks throughout the life of a program (especially in the early stages) to assure the quality of the data. Evaluators should discuss their data requirements with the corporate information manager before the program or the evaluation gets underway rather than after it is started.

STAFFING AND BUDGETING

Once general evaluation planning has started, the utility will need to consider both budgeting and staffing. Clearly, the evaluation effort should be commensurate with the scope of the program, the importance of the program, and the use to which the evaluation results will be put. The cost of an evaluation can range from 1 to 10 percent of DSM spending. Staffing may vary greatly, depending on corporate structure and how consultants are used.

Staffing

An early consideration is where evaluation fits in the corporate structure. Utilities handle this differently, and there is no right answer. Typical structures include:

- Evaluation and implementation can be kept totally separate (e.g., under a different vice president) as is currently the case at Central Maine Power (CMP)
- Evaluation and implementation can be placed under the same vice president with different directors or managers, as is currently the case with New England Electric System (NEES), Pacific Gas & Electric (PG&E), and Northeast Utilities
- Evaluation can be combined with program planning with or without the same vice president as implementation
- Evaluation can be combined with load research or market research
- Evaluation can be combined with load forecasting

Each of these structures has its strengths and weaknesses. Separating evaluation from implementation avoids perceptions of conflict of interest and is effective provided frequent and good communication occurs between the groups. Having evaluation and implementation under the same vice president can ensure that conflicts are resolved quickly but may lead to perceptions of a conflict of interest. This arrangement also increases the likelihood that evaluation results will be used in revising and planning programs. Combining evaluation with planning has the advantage of sharing personnel and skills but may lead to situations at startup in which personnel are more focused on planning and refining programs than on evaluating programs (Franklin and Thrasher 1976). Combining evaluation with load research can lead to strong data collection. A combination with market research may provide economies of scale in company-wide market research as well as provide evaluators with good access to customer information; but it may tend toward a focus on market issues rather than technological or persistence issues. It is hoped that a utility will structure evaluation activities to achieve the strongest possible balance among competing interests.

Another consideration concerning staffing is the amount and type of evaluation work to be done by in-house personnel and by consultants. This decision will determine the type of personnel that the in-house evaluation group will need. For reasons of credibility and objectivity, utilities generally prefer to have consultants perform process evaluations. This is not to say that a distinctly separate in-house evaluation group could not undertake process evaluations. Indeed, when evaluation and implementation had different vice presidents at NEES, some process evaluations were done in-house. Even in this situation, though, a utility might use a consulting firm to conduct survey work unless the company has a strong market-research department capable of doing its own surveys and having its own phone bank for surveys. Most utilities will probably opt to contract for process-evaluation work.

Impact evaluations can involve analysis of customer electricity bills, metering studies, and engineering studies. These can be done by in-house staff or by consultants. Billing and other statistical analysis of customer kWh data typically involves sample design, extraction of customer information from implementation databases, extraction of billing data from billing files, customer surveys, and the analysis of bills. Even if a consultant is used, the utility staff will probably have to extract the data from the implementation and billing databases, a time-consuming step. The decision to complete the other tasks with in-house personnel should be based on the qualifications and availability of utility staff. A utility with strong load-research capabilities and a strong metering department may find it easier and less expensive to do its own metering. Significant advantages accrue to having utility staff who understand metering devices and are available to maintain metering equipment. In some cases, a utility may want a mix of direct employees and contract workers: a utility might use its own staff to place end-use meters on residential water heaters and to place whole-building meters at SH-customers' sites, and might use a contractor to install end-use meters on HVAC systems. Engineering analysis could be done by utility staff or by a contractor.

In-house personnel requirements are determined by the decision on how much work will be done by consultants. If most evaluation work is being done by consultants, the utility need not have specialists in every evaluation area. In this situation, utility staff should probably be good project managers and have experience managing contracts. As discussed in Chapter 1, appropriate backgrounds include market research, economics, psychology, sociology, policy analysis, statistics, and engineering. If the utility chooses to do much impact-evaluation work in-house, expertise in certain areas is essential. These areas include econometrics or statistics for statistical and billing analysis, engineering analysis, and electrical and electronics engineering for load-research and metering studies.

Evaluation staff need to have good analytical and communication skills. The ability to write effectively is important for presenting evaluation results in a usable format. Computer skills, particularly knowledge of spreadsheet and statistical packages, are essential. Good interpersonal skills are needed to be able to interact with people from other departments.

Budgeting

Budgets will depend on the number of programs being evaluated, the available staff, the amount of work being done by consultants, and whether some work is being done by other departments. Often, in the first year of a program, only a process evaluation will be conducted because it is too early to use impact-evaluation techniques.

Table 5 provides crude guidelines for budget-estimation purposes. Clearly, the utility will need to follow its own procedures in determining a budget. Metering costs are not included, and they will vary greatly depending on the type and duration of metering and whether metering is done by utility staff or by a consultant. The load-research and metering staffs at the utility can provide good estimates of metering costs; (see also Table 22 in Chapter 7).

Table 5. **Typical evaluation costs**

Item	Person-hours	Approximate cost @ $90/hour (1991-$)
Process evaluation		
Residential program	300 – 800	27,000 – 72,000
C&I program	600 –1,500	54,000 – 135,000
Impact evaluation, billing analysis		
Residential program	400 – 700	36,000 – 63,000
C&I program	600 –1,800	54,000 – 162,000

COORDINATION WITHIN THE UTILITY

Regardless of the structure of the utility, evaluators must be prepared to work closely with the following departments: program implementation (field staff), load research, information systems, metering, program planners, load forecasting, market research, and resource planning. Coordination with program implementers is vital for producing usable and reliable evaluation results. Evaluators should be involved with program planning, and implementers should be drawn into the evaluation process early and provide continuing feedback. The load-research and metering departments are critical if the company plans to do any metering on its own. Selection of metering equipment and protocols needs to be consistent with that for any existing load-research system. Such issues as additional workload for union employees may need to be addressed with the metering department. Interaction with the information-systems department is very important. If the information needed by evaluators is collected at the outset, subsequent tasks will be easier. The load-forecasting and market-research departments may have baseline data that can be used by evaluators. In turn, evaluators can provide these departments with additional information on customers.

SELECTING AN APPROPRIATE EVALUATION APPROACH

The selection of an evaluation approach for a given DSM program depends upon the size and the life-cycle stage of the program, the available resources, the goals of the program, the goals of the evaluation, how the evaluation will be used, regulatory interest, the possibility of joint efforts with other utilities, and whether multiple evaluation approaches might be appropriate.

Process evaluations usually examine participation, nonparticipation, operational efficiency, contractor-utility interactions, usefulness of the corporate and implementation databases, effectiveness of marketing, customer satisfaction, implementation effectiveness, and free ridership. The methods used in process evaluations usually consist of customer surveys, interviews with appropriate staff and contractors, and field visits to locations where measures are installed. However, the foci of the process evaluations tend to vary. Evaluations of new programs often focus on customer satisfaction with the program and suggestions for program improvements. If free ridership or program delivery are of concern, these issues may be examined in more detail (e.g., by asking retailers their impressions about changes in consumer buying habits or by comparing sales figures for specific models of appliances before and after a program has started). The scope of the process evaluation for a given program must be based on a utility's assessments of its needs, concerns of regulators and other parties, and the resources available for the evaluation.

The choice of impact-evaluation method may be influenced by the use to which the results of the impact evaluation will be put (will it be used for regulatory incentives?), the stage of the program (is this a new program and how many participants does it have?), the costs and size of the program, the utility operating characteristics, and regulatory interest. More rigorous approaches and multiple methods are appropriate when the utility receives money (e.g., DSM incentives for shareholders or wheeling of conservation) based on the evaluation results; where large resources are being expended on a given program; and, in R&D programs, where the evaluation results will be used to develop a larger-scale program.

As an example, consider two utilities with weatherization programs aimed at reducing electricity use for SH. The first utility has an immediate capacity need. Its peak is driven by heating, it receives incentives based on evaluation results, and it is spending about 1% of its annual revenues on this program. The second utility does not need new capacity until the year 2005. It is summer peaking, it just receives cost recovery on its conservation programs, and it is spending 0.1% of its revenues on this program. Both utilities might perform process evaluations, and they also might estimate kWh savings through billing analysis. The first utility also might use end-use metering to evaluate the capacity benefits from the program and to verify both kW and kWh savings because incentives are involved. The second utility may not use end-use metering and use those evaluation dollars elsewhere because it has no immediate capacity requirements, is summer peaking, and is offering no incentives.

Evaluators should remember that, for some programs (e.g., a low-flow showerhead give-away program), accurate estimates can be obtained with simple techniques (e.g., engineering estimates) and that the use of more-sophisticated and -costly techniques may not buy corresponding increases in accuracy (EPRI 1991a).

Certain types of programs lend themselves to combined efforts with other utilities. An example might be a program to develop baseline savings estimates for new construction practices. Four Massachusetts utilities used this approach (XENERGY 1991; New England Power Service Company 1991). A combination program of metering and statistical estimation of load-shape impacts has been proposed for New York state (XENERGY 1990). Several joint evaluations have been conducted in Wisconsin (Prahl 1988). Also, several utilities in New England are performing a joint evaluation of an appliance-efficiency program.

RESEARCH DESIGN

The savings from a DSM program are estimated by comparing energy or demand prior to a program with energy or demand after a program has been implemented. Figure 6 is a conceptual representation of this process. One approach is to compare the posttreatment consumption in year three (Point C) with preimplementation consumption (Point A). The difference (A – C) represents the effects of the program. However, this difference also includes the effects of all the other changes that occurred in the intervening three years. If this customer were a commercial building, this difference might include changes in occupancy in the building; increased energy-use intensity caused by the introduction of new equipment, such as laser printers; increased or decreased activity in the building because of changes in the economy; etc. If it were a residential dwelling, a variety of factors might affect energy or demand, including changes in the composition of the household (e.g., the addition of a baby or a young adult leaving home); the introduction of new appliances; the replacement of older appliances that were not part of the program with new more-energy-efficient appliances; or changes in lifestyle, such as the maturation of young people.

Thus, to accurately estimate the savings of a DSM program, one needs to measure the difference between the electricity that would have been consumed over time without the program and what is actually consumed with the program. Figure 6 shows that without the

Table 6. **Examples of threats to internal validity**

Threat	Example
History	Another event such as an economic crisis, a war, or an "oil shock" causes a change in electricity consumption.
Maturation	Natural changes, such as an addition to a family, component aging, or changes in lifestyle, occur which cause changes in energy consumption.
Regression to the mean	Buildings or households chosen for an initial measurement from the extremes of a distribution tend to have measurements nearer the middle of the distribution on subsequent measurements. If high energy users are targeted for the program and measured in the initial interval, high energy use may not be sustained in subsequent measurements.
Selection	Selection is the situation in which participants are systematically different than the remainder of the population. Selection bias is often a problem with voluntary programs.
Testing	People adjust their behavior or adapt to changed circumstances on the basis of programmatic activities. People may adjust their thermostats in response to AC load-control tests. People "take back" some of the savings from a weatherization program.
Mortality	Participants drop out for some reason such as relocation or dissatisfaction with the program.
Instrumentation	Changes in how the impact variables are measured account for the changes. A question on a survey is changed, which causes people to answer the question differently because they assume a different context, or metering is changed from whole-building metering to end-use metering between the pre- and postperiods.

program than would the general population. Thus, if projections of reductions in energy usage from people volunteering for an advanced-heat-pump program includes a disproportionate number of innovators, the total amount of savings may be overestimated.

Campbell and Stanley (1966) argued that the "classical control group" design minimizes these threats. In this design, pre- and posttreatment usages by a group of participants and a group of nonparticipants are compared. The key to the classical design is the *random* assignment of customers to the participant and nonparticipant groups. Randomly assigning customers

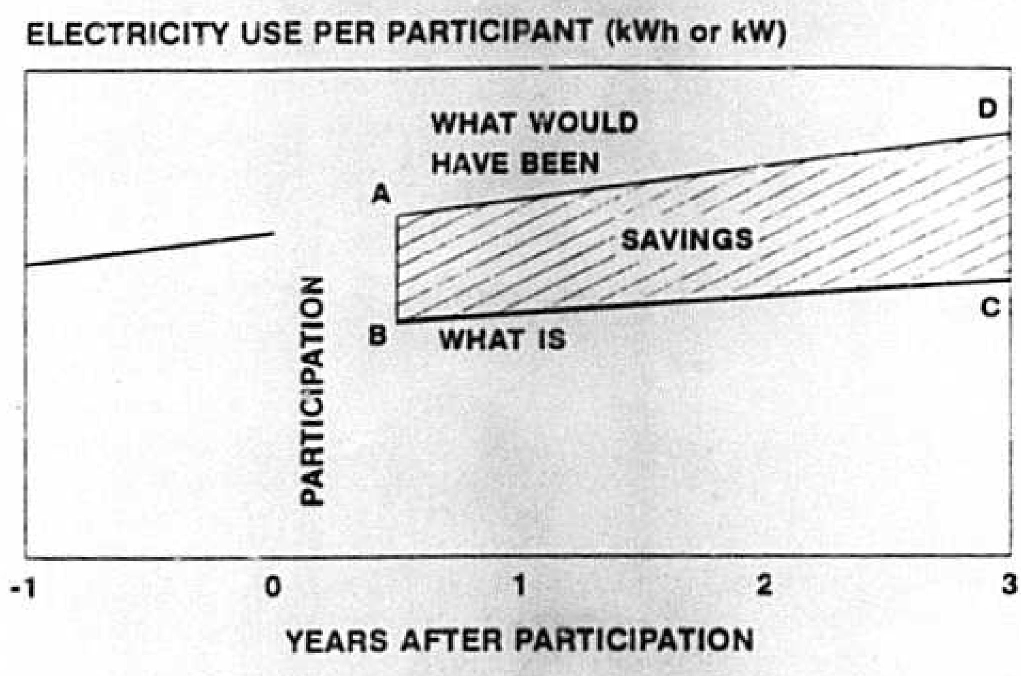


Fig. 6. Schematic showing electricity use before participation in a DSM program, after participation in the program (BC), and an estimate of what electricity use would have been without the program (AD). The difference A – B is the immediate postparticipation change in electricity use. The difference D – C is the long-term change in electricity use.

program, demand or consumption increased during the three-year period, the line with A and D as endpoints. Even with the DSM measures, energy or demand increased some, as represented by the line BC. Thus, the savings are greater than would be represented by the quantity A – C and would perhaps be better represented by the quantity, D – C. The cumulative savings are represented by the area, ABCD.

The purpose of research design is to minimize the impacts of extraneous effects of the type described above. These extraneous factors are rival hypotheses or "threats to internal validity." Threats to internal validity arise when the differences that exist can be explained in part or in full by other phenomena. Cook and Campbell (1979) list seven potential threats to internal validity (Table 6) that, if not controlled by the design, can bias results.

In addition to internal threats to validity, there are external threats to validity. External threats are those factors that make generalization of findings difficult. An example would be findings from a sample of customers that is not representative of the larger population of interest. Another example is a reactive or interactive effect of testing (e.g., where pretests of marketing materials during a pilot program lead to a decreased sensitivity to the materials during a full-scale program). Yet another form of external threats to validity is the interactive effect between selection biases and the experimental variable. In this situation, customers who are innovators may be more inclined to participate in a technological to both groups reduces the likelihood that these groups will be biased in some way or that the results will not be generalizable. Also, in the strictest sense, participants and nonparticipants should not know to which treatment group they are assigned and nonparticipants should be isolated from treatment-related information received by participants. With the classical design, the change in energy usage is estimated as follows:

Change in usage = (participant pretreatment usage - participant posttreatment usage) - (nonparticipant pretreatment usage - nonparticipant posttreatment usage)

Unfortunately, practical circumstances almost always preclude the use of the classical design. Examples of such circumstances might include a PUC order that a program must be offered to all customers or an order that the same incentives be available to all customers. Each of these orders would preclude the random assignment of customers. Yet another case would be a rebate program in which it is impossible to treat some customers without influencing others.

In these circumstances, the evaluator must look to alternative designs to minimize, but not necessarily eliminate, threats to validity. Table 7 presents some of these options. For these designs, the nonparticipant group is called a comparison group because nonparticipants are not randomly assigned (as a "control group" would be) and thus are not strictly equivalent.

Table 7. **Research-design options**

Design option	Pretreatment period	Participation	Posttreatment period
Classical design (random assignment)			
Participants	0	X	0
Nonparticipants	0	--	0
Nonequivalent comparison group			
Participants	0	X	0
Nonparticipants	0	--	0
Cross-sectional comparison group			
Participants	--	X	0
Nonparticipants	--	--	0
Time-series comparison participants	0	X	0

0 = observation; X = participation; -- = no data.

However, this does not mean that participants and nonparticipants may not have been randomly selected.

The nonequivalent comparison group design tries to make up for the absence of random assignment of participants and nonparticipants to treatment groups by constructing a comparison group that is similar to the participant group. This fashioning of the comparison group may be done by matching the characteristics of nonparticipants with selected characteristics of participants or by drawing a random sample weighted for a characteristic (or group of characteristics) from the population represented by the participants. An example of this approach might be to compare those who volunteered for a program with a weighted sample of customers based on total kWh sales during a pretreatment period. If the two groups are comparable, then threats to internal validity may be minimized, but external validity threats (i.e., the generalizability to the larger population) may remain.

In a design with a cross-section comparison group, the energy usages of two groups (participants and nonparticipants) are compared for the posttreatment period only. If the selection of participants is random or the two groups are well matched, then many of the threats to internal validity are minimized, although there still may be problems with generalization. Often, this design is used in situations where an evaluation is initiated after a program has begun or where the utility does not have historical billing data for a group of customers or where no pretreatment period is available (as in new construction). If evaluation planning is part of the program planning, problems of obtaining pretreatment data will be minimized.

Another design, a time-series comparison group uses the participants as their own comparison. For example, in a load-control program, electricity use on noncontrol days with similar weather is often used as the baseline against which to compare electricity use on control days. Figure 14 in Chapter 7 illustrates this approach for a residential air-conditioner control program. For such programs, this time-series comparison is a strong design because the short time span and the use of weather conditions to select noncontrol days preclude most threats to validity. This type of comparison group might apply to other programs as well. If the participants are randomly selected, then threats to internal validity, with the possible exception of history and perhaps instrumentation, are minimized. Many of these problems can be dealt with by the use of normalization techniques or the use of multivariate analysis. If participants are not randomly selected, then many of the threats to validity remain.

It should be clear from the preceding discussion of research designs that sample design is a critical issue. The lack of a good sample design can pose threats to both internal and external validity. This subject will be discussed in some detail in the next chapter.

CONCLUSIONS

Planning an evaluation involves an array of activities. Many of these activities revolve around identifying the goals of the evaluation and obtaining agreement on the goals from stakeholders. Agreement on goals is essential for obtaining commitment to the evaluation and for ensuring that evaluation results are used. A clear set of evaluation goals are also essential to ensure that the appropriate data are collected and available for analysis when needed.

The choice of evaluation approach depends on the type of program, size of program, customer segment, and ultimate use of the evaluation's results. Finally, the choice of a research design is an important part of the evaluation plan. A good research design will guide the where, when, and how of data collection and it will help ensure that the results can withstand the challenges of alternative explanations of the findings.

MONITORING THE IMPACTS OF DSM PROGRAMS

Harvey Michaels
Marc Hoffman
Andrew Schön

INTRODUCTION

This chapter addresses monitoring and data collection. It's objective is to describe the major options available to the DSM-program evaluator for data collection, data specification, and sample design and to provide an assessment of the strengths, weaknesses, resource requirements, and implementation approaches for the various data-gathering techniques.

Monitoring is the process of collecting the data needed for the analysis of energy savings, including identifying the data requirements to establish energy savings, normalizing energy use for nonprogram factors, estimating free-rider impacts, and determining persistence of savings. Monitoring includes data specification, survey design, sample selection, metering installation, data collection, quality control, and reporting.

The selection of data-collection method, data specification, and sample design are key decisions that must be made jointly by those responsible for planning, monitoring, and analysis. These decisions determine the cost of the evaluation and the quality of the analyses that can be performed. This chapter discusses the objectives of monitoring, the strengths and weaknesses of various sources of monitoring data, how to specify the data to be collected, sample selection, and data-quality issues.

OBJECTIVES OF MONITORING

The primary objective of monitoring is to provide systematic, defensible, and objective measurements for a DSM-program evaluation. Monitoring, therefore, should be carried out in a manner that achieves a quality research standard, such as the following standards identified for educational evaluations (Joint Committee on Standards for Educational Evaluation 1981, partial and edited; Patton 1982a):

> The purposes and procedures of the evaluation should be monitored and described in enough detail so that they can be identified and assessed.
>
> The sources of information should be described in enough detail so that the adequacy of the information can be assessed.

The information-gathering instruments and procedures should be chosen or developed and then implemented in ways that will assure that the interpretation arrived at is valid and reliable for the given use.

The data ... should be reviewed and corrected so that the results will not be flawed.

The evaluation procedures should provide safeguards to protect ... against distortion by the personal feelings and biases of any party to the evaluation.

WHAT DATA TO COLLECT

The purposes of data collection are to establish the quantity of energy and peak-demand savings and to support comparisons, projections, and program design. Almost all impact evaluations seek to quantify kW and kWh saved. Data collected to support these calculations can include billing data and metering data as well as inputs to engineering or statistical models. Data of this type is often the highest priority because of its relationship to utility rates, profits, and/or cost recovery.

As was pointed out in the previous chapter, external validity is important for generalizing conclusions. To perform such comparisons and projections, additional data (such as customer location, market-segment information, customer preferences, weather, and economic variables) are needed. If these data are collected, then impacts may be normalized, making them transferable, which enhances the likelihood that meaningful comparisons can be made with similar programs in other service territories. The collection of this type of data also makes it possible to develop predictive models, which can be used to estimate program performance for different service territories, customer segments, and economic conditions. Being able to posit causal relationships opens the possibility of projecting impacts for programs with different designs and of reassessing program impacts in response to changes in avoided costs, the economy, or utility or regulatory priorities. It also supports program redesign to improve future performance.

The previous chapter pointed out that the demand or energy savings are the differences between what would have been consumed without the program and what is actually consumed with the program. The data requirements for DSM-impact evaluation are

- The usage and demand of the facility prior to measure installation and changes in operating characteristics of the end-use or facility
- The daily or annual operating profile of the energy-consuming systems affected by the measure
- The operation and maintenance (if applicable) of the measure and the life of the measure (persistence), including estimates of the decline in efficiency of the measure and/or its removal

- The data to estimate what would have been had no program been offered

- The net change in power or energy requirements resulting from the measure

Because measurements and estimates contribute to the savings calculation, measurement error in any component will be carried through to the assessment of the impact. Making measurements that are rigorous for one or two elements will not result in an accurate estimate of impacts if the remaining parameters are not measured or are measured in a limited way. Accuracy of the total evaluation is limited by the accuracy of the weakest measurement. Therefore, the ideal approach is to develop a balanced data-collection strategy, where variance for each measurement is about equal.

The requirements for data collection are determined primarily by the analysis method chosen (Table 8). There are three general types of analysis methods and associated data:

- Direct measurements, which are used to calculate changes in energy use by comparing measurements made at different times. Direct measurements include customer billing, whole-building metering, end-use metering, and frequency meters.

- Engineering modeling, which uses physical models ranging from simple tracking models to complex physical simulations of buildings to analyze energy use. Data used in these models include weather, customer surveys, facility and equipment inventories, and operating patterns.

- Statistical modeling, which uses statistical techniques to evaluate changes. Data that are often used in these types of models include billing data, market-segment information, demographic characteristics, and economic variables, such as electricity price.

These techniques are often used in combination. Billing data are often combined with customer data for analysis. For example, a regression of monthly energy use against weather data can be used to detect the presence of heating, cooling, and nonweather-sensitive uses as well as to determine individual use levels. Conditional-demand analysis uses direct measurement and statistical regression (Chapter 4). Hybrid statistical/engineering method (HSEM) refers to the use of engineering and direct-measurement data to drive engineering based statistical regressions.

Based on the objectives and analytic methods, the evaluator can define procedures and specifications for data collection. These specifications should include the source, the type (e.g., definition of units and tolerances), sample sizes, frequency of data collection, and quality-control requirements.

Table 8. Data options by analysis method

Data source	Analysis method			Combinations		
	Direct measurement	Engineering model	Statistical model	Billing analysis	Conditional demand	HSEM
End-use metering	■					
Whole-building metering	■				□	□
Billing data	■			□	□	□
Facility inspection		□				□
Mail survey			■		□	
Telephone survey			■		□	
Onsite survey			■		□	
Weather data		□		□	□	□

■ Data source sufficient for analysis method.
□ Additional data source required.

TYPES OF INFORMATION FOR IMPACT EVALUATIONS

The variety of data sources includes facility energy-use measurements, facility inspections, customer-supplied data, research to establish typical values, and combinations of data sources. Determining which source to use depends on the evaluation objective, analytic method, costs of data collection and analysis, and accuracy. There is no best or worst source; each has strengths and weaknesses (Table 9).

Facility Energy-Use Measurement

Facility energy-use measurements can come from customer electric and fuel bills, hourly consumption from whole-building meters, and end-use metering. These are direct measurements and should be accurate within the tolerances of the metering device. The evaluator needs to remember that these are measurements of present use and not changes in use; they have no explanatory power in and of themselves.

End-use metering is expensive but has the advantage of providing a direct measurement of load for the device or circuit of interest. With advances in solid-state technology and experience, the costs of end-use metering are dropping. Recent costs for the metering device, installation, data-acquisition system, and maintenance range from $5,000 to $20,000 (in 1991 dollars) per building per year; see also Table 22 in Chapter 7 and Exhibit 4. Several projects conducted in the past five years have sampled three to eight loads in each of 25 to 100 buildings.

Table 9. Characteristics of data sources

Data source	Strength	Weakness	Cost
1. Facility energy-use measurement	Accurate to bounds of the meter	Measures present use not change parameter; no explanatory power	Varies, see below
• End-use metering	Highest unit accuracy of the target system	Most expensive per unit; budget limitations result in high sampling error	$5,000 to $20,000 per building
• Whole-building metering	Measures short-internal whole-building loads that can indicate target-system loads	Lower unit accuracy than end-use metering	$500 to $3,000 per building
• Customer billing data	Census eliminates sampling error	Difficult to detect change in treated system	$2 to $10 per building
2. Facility inspection	Collects causal technical factors best	Time limitations at site cause many values to be estimated	$20,000 to $40,000 fixed costs plus $100 to $600 per facility
3. Information supplied by customer	Collects customer market data and attitudes best	Error rates high on technical data	See below
• Mail survey	Low cost permits large sample	High error rate and risk of nonresponse bias	$10,000 to $20,000 fixed costs plus $15 to $40 per survey
• Telephone survey	Higher response rate and lower error rate than mail surveys	Higher cost than mail surveys	$15,000 to $30,000 fixed costs plus $25 to $60 per survey
• In-person interview	Lowest bias and error on customer questions	Higher cost	$15,000 to $30,000 fixed costs plus $150 to $300 per site

Exhibit 4. The costs of multichannel, automated, end-use monitoring

Components of this cost are highly variable, but recent ranges for a full-scale data-acquisition system with a central personal computer (PC), remote meters/recorders, and communication system are as follows.

- Meter and wiring costs depend on the number of channels, how the monitored data is communicated to the building recorder, and the difficulty of isolating the monitoring point from the rest of the electrical system. Costs for a building recorder and meter transponders are $500 to $2,000 per facility while specification and installation add an additional $500 (residential) to $3,000 (complex commercial) per facility.

- Data-acquisition system set-up costs range from $20,000 to $100,000. In addition, the annual costs per facility for data acquisition and management are $500 to $1,000.

- In total, end-use-metering fixed costs range from $20,000 to $100,000 for the first year plus $1,500 to $6,000 per facility annually.

- Analysis of the data bears additional costs.

Barring technical problems with the wiring, the meter, or the data-acquisition system, the electronic end-use meter provides a high-resolution measurement (frequent recording of values), which is often the standard for accuracy to which other measurements are compared. End-use metering is the only measurement technology that combines high resolution, accuracy, and isolation of the energy-consuming system. As a selective research tool, it can be applied to particular questions that other data-collection forms fail to answer satisfactorily. For example, end-user monitoring may be the only technology that can be used to accurately resolve the effect of lighting efficiency on cooling-energy use, an important issue in lighting programs.

Because of their cost, end-use meters tend to be used with small samples, which means that extrapolations to the total population produce a fairly large error. Several studies, notably the evaluation plans of Northeast Utilities and Empire State Energy Research Corp., have found that other methods may produce more-precise population estimates because they use larger sample sizes at lower costs (Townsley and Wright 1990; XENERGY 1990). Another possible problem with end-use metering is that the customer may be aware of the monitoring and may change behavior (the Hawthorne effect).

In addition to the multichannel, automated, end-use metering described above, other types of end-use monitoring exist, many of which have lower costs. These methods include instantaneous metering, portable/temporary electronic metering, and nonelectronic portable/temporary metering.

Instantaneously recorded amperage, voltage, wattage, power factor, frequency, etc. on a single appliance can be used to establish instantaneous load conditions, such as the kW for air conditioners, lighting fixtures, motors, heating elements, etc. A variety of handheld and portable meters are available to measure these values at costs ranging from less than $100 to $5,000. These devices are useful for providing field data to validate engineering estimates. In the case of a pure unswitched lighting circuit, this type of monitoring device might be used to evaluate net change in load by determining connected load before and after high-efficiency lighting is installed. In this example, care needs to be taken to account for burned-out bulbs and fixtures. Generally these meters are not useful for gathering data where the interest is in daily, weekly, or seasonal variations.

Portable/temporary meters can function over a longer period, ranging from several days to several months. Usually these systems take measurements, store the measurements in a recorder, and subsequently upload the data to a PC for analysis. These systems cost between $500 and $700 per recorder channel and come with meter/recorder hardware as well as PC software for analysis, graphics, and reporting. With portable/temporary meters, daily, weekly, or seasonal variations can be captured without communication lines.

Portable nonelectronic meters are similar to electronic meters except that the data are collected on circular/strip charts, cumulative value registers, or runtime meters. Utilities often overlook the fact that the simplest and cheapest form of end-use metering can be a standard watt-hour meter, which can be connected to an end-use load. This works as long as the measurement of interest is cumulative kWh and some provision can be made for reading the meters at the desired frequency.

Whole-building load-meters are used to record building energy use at short intervals, such as each 15 or 30 minutes but sometimes as frequently as each minute. These meters are less expensive than end-use meters and are often installed without the knowledge or consent of the customer. The cost of the meters is similar to that of end-use meters, but installation (usually $200 to $500), the data-acquisition system, and annual meter management are less expensive. Because many utilities already use such meters for load research and routine metering of large customers, the fixed costs may be reduced. Depending on which of these circumstances apply, total load metering can cost $500 to $3,000 per facility.

While providing an accurate measurement of total load, whole-building meters are at a disadvantage compared with end-use metering for measuring subsystem loads. They do not provide a direct measure of the treated system. If the treated system is a large fraction of the total load, the change in load from the treatment is large, and the frequency of measurement is high, accurate estimates of end-use loads and changes in end-use loads can be extracted. For example, HVAC loads, which are responsive to weather, are often a statistically significant component of the load and can therefore be estimated with whole-building meters.

The lower cost of whole-building meters permits larger samples and may increase population accuracy for a fixed evaluation budget. Whole-building meters are being assessed as a primary data source for HSEM. This form of analysis, which also requires a facility inspection, is being investigated for comparative bias and accuracy to end-use metering and other methods (XENERGY 1990).

A powerful extension of the whole-building metering concept is the appliance signature or nonintrusive meter, which records variations in real and reactive load components. As appliances turn on and off, the signature of these components is identified and recorded electronically. This concept has been tested in residential applications, and it explains up to 80% of the load (EPRI 1989).

Standard customer billing data are available for all participants and nonparticipants. Those data can provide such information as monthly or bimonthly kWh consumption, monthly peak demand for large commercial accounts, and (occasionally) time-of-use measurement. Because these data are collected by utility billing operations and are generally maintained on customer-information databases, the acquisition costs are very low. The ability to use a census gives the analysis of this data the potential for eliminating extrapolation error. The disadvantage is that the limited information on each customer makes the detection of changes in consumption for individual systems difficult. The energy consumption of loads that are large and fluctuate with the season can be differentiated. The Princeton Scorekeeping Method (PRISM; see Chapter 4) uses billing and weather data along with a statistical heuristic to produce end-use-specific estimates of energy use.

Facility Inspection

Facility inspections collect physically observable data, such as nameplate data and information about observable operating conditions, through energy audits and site surveys (Fig. 7). These data permit independent verification of customer- or implementor-supplied information and provide greater detail and consistency in the collection of observable parameters. Facility inspection data often are used for the detailed inputs to engineering models.

The reliability and accuracy of predicted energy consumption and hourly load estimates from engineering models is limited by the difficulties of collecting the data that serve as model inputs. Many of the data requirements for engineering models are related to the behavior of the building occupants. Accurately describing occupant behavior requires extensive observations, so it is usually estimated based on a brief facility inspection and reported behaviors. For example, the occupancy profiles and operating schedules of equipment may be observed during a facility inspection, but data about off-peak occupancy and operating schedules may be based on the customer's statements and perceptions.

The amount of time available for inspection may limit the number of measurements at a site with attendant reductions in the overall quality of facility-inspection data. For example, in inspections of large or multibuilding facilities, building attributes such as envelope thermal characteristics, equipment efficiencies, lighting intensities, ventilation rates, etc. are generally sampled or estimated to reduce costs. Then the results are extrapolated to the whole facility.

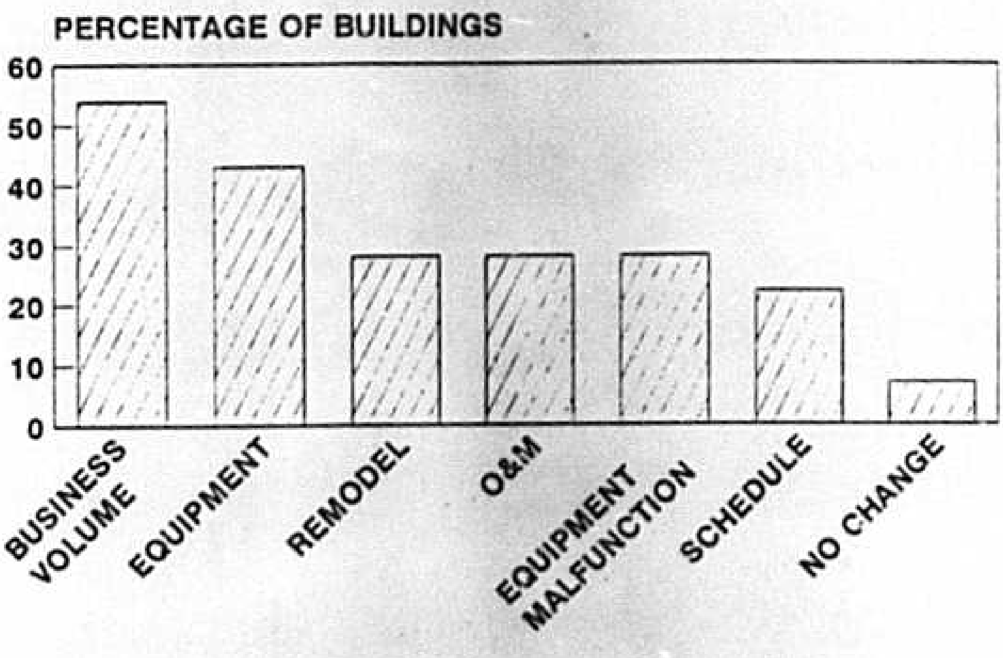


Fig. 7. **BPA conducted onsite inspections of 46 buildings one to six years after the facilities had participated in its Institutional Buildings Program. These inspections found many changes that probably affect postretrofit electricity use.**

A brief facility inspection does permit collection of most of the technical data required for modeling a building. By limiting observation, homes and small commercial facilities can be inspected in as little as 30 minutes, medium to large commercial facilities in 1 to 4 hours, and very large commercial or industrial facilities in a day or less. By way of contrast, a complete engineering inspection that includes measurement and/or observation of all significant engineering factors can take two to five times as long.

The total cost of an engineering facility inspection includes scheduling and travel time, data preparation, data entry, and quality-control procedures. Typical fixed costs are $20,000 to $40,000 to prepare the survey instrument and protocols, recruit and train inspectors, field test the instrument and survey process, and develop quality-control procedures. It is difficult to complete more than two surveys per day even if the average inspection time is only an hour. Because more than half of the typical inspector's time is associated with travel, schedule limitations, waiting, and canceled appointments, considerable savings are associated with inspections conducted during visits to the customer for other purposes. Unit costs are $100 to $300 if the inspection is part of an otherwise scheduled visit or $250 to $600 if conducted separately.

Customer-Supplied Information

Information is often supplied by the customer in response to a survey rather than measured or observed. Information to determine the customer's market segment, equipment-operating patterns, actions, and attitudes can be collected as part of the application for program participation or by response to mail, telephone, or in-person interviews.

Using customer-supplied data has advantages. Data from this source are inexpensive compared with data from other sources. And this is the only method for gathering direct input on customer attitudes, opinions, and preferences.

However, customer-supplied data has some disadvantages. Customers cannot be relied upon to supply detailed technical data. Customers may erroneously report data about the physical and operating characteristics of a facility or appliance. Customers may not know the answers. They may not be motivated to determine the answer. Or, they may give false or misleading answers. It is not uncommon for customers to confuse a hydronic heating system with the domestic hot water system (they are both water-heating systems). A customer is unlikely to know the loading on motors. Customers cannot be expected to respond to burdensome questions, such as asking a business owner to report the nameplate sizing of AC units or the number of four-lamp fixtures that currently have two lamps. Reports may be biased by the desire to report a "tight ship," for example, by indicating that the lights are off when the facility is unoccupied. Appraising customers' willingness to purchase equipment or determine future operating characteristics of equipment based on questionnaires often fails because customers' attitudes do not predict their behavior (Converse and Presser 1986).

The three basic methods of soliciting customer information are the mail survey, the telephone interview, and the onsite interview. Mail surveys are inexpensive, costing $15 to $40 per completed survey, including a computer-generated report. The size of the sampled population is the greatest determinant of costs. Survey design and protocol development are typically $10,000 to $20,000. Mail surveys often have low response rates and high error rates. Response rates for mail surveys can be 5 to 10% in commercial surveys and 20 to 40% for residential surveys. High error rates result from a lack of incentive to complete the survey, carelessness, ambiguity, and lack of supervision of the respondent. The problems can be mitigated by improved packaging of the questionnaire, better timing, providing a clear and nonjudgmental set of questions, and providing an incentive to respond (Dillman 1978; Train 1988).

Telephone surveys cost more than mail surveys. Survey design and protocol development tend to be $5,000 to $10,000 higher than for mail surveys. Conducting the survey costs $25 to $60 per completed questionnaire. Telephone surveys generally have higher response rates and lower error rates than mail surveys. However, with the increased use of telemarketing by businesses, the public is becoming increasingly wary of unsolicited calls, and researchers are reporting difficulties in getting an adequate sample. Telephone response rates can be improved by sending a letter in advance of the call explaining the purpose of the call in general terms.

Respondents to a telephone survey cannot be expected to provide specific physical data about equipment and often can provide only a very general indication of the types of appliances they have. For example, residential customers may not be able to distinguish between heat pumps and strip heaters unless survey questions are carefully worded. Telephone surveys can be worded so that they provide an indication of types of equipment, satisfaction with performance, type of facility, operating hours, and attitudes (Dillman 1978).

An in-person interview allows for more depth with the obvious trade-off against cost. If conducted in isolation, interviews cost $150 to $300 per interview. If facility inspections are conducted for physical data, the marginal cost of an interview is usually small, perhaps an additional 30 minutes of site time. The costs associated with survey design and protocol development are similar to those for telephone surveys. Unstructured interviews can be very useful in eliciting an understanding of a customer's or project manager's experiences, especially in the initial stages of a project. However, unstructured interviews are difficult to analyze. The analysis of unstructured interviews requires the use of content-analysis techniques, and the results are difficult to generalize. Open-ended questions in structured interviews have many of the same problems. Close-ended questions have the virtue of asking the same question and providing the same response set to each respondent. This makes the responses easier to analyze and interpret, but if the questions are not well designed, the information obtained can be limited.

Research to Establish Typical Values

Bench testing of equipment or borrowing data from other utilities are two ways that utilities can establish typical values. Many impact characteristics can be established by bench testing based on data provided by manufacturers or independent laboratories. A laboratory measurement of unit savings in combination with reported information on installation, operating patterns, and prior conditions can be used as inputs to engineering models.

This approach is not without risks. The conditions under which a manufacturer's measurements are made may not be typical of conditions in the field. Further, conditions in the field may interact with the equipment to change the operating characteristics. The operating patterns used in tests may be based on extreme patterns or those that are assumed to be typical but have not been validated.

Facility operating patterns, baseline characteristics, customer maintenance behaviors, and persistence may be available from a similar utility or may be predicted by examining nonenergy products with similar economics for the customer. A literature review provides an important starting point for predicting typical values.

Collecting Local-Weather, Equipment, and Customer-Characteristics Data

Economic, demographic, and weather data are almost always essential for establishing a baseline. Sources for these types of information include the population census; the housing census; special censuses, such as those for C&I establishments; economic statistics; and EIA's

Commercial Buildings Energy Consumption Surveys. Private databases on construction, space inventories, and equipment sales also may be useful. A utility may maintain records of new construction notifications or permits (Weiss 1972).

Combinations of Data Sources

Evaluations often use combinations of data sources to measure impacts (as discussed in the section on Selected Special Issues in Chapter 4). This practice has significant implications for data-collection procedures. Triangulation is the term used to describe the parallel use of multiple data sources. A common approach is to use billing analysis for the population or a large sample, end-use metering for a small sample, and a field survey and engineering models for a medium-sized sample. This may result in three different impact outcomes. A triangulation strategy uses combinations of data to improve the overall estimate. For example, the metered sample can be used to determine the bias in the engineering approach. It can also, using Bayesian mathematics, be used to reduce the variance in the billing analysis.

Samples used in triangulation can be independent of each other although there may be some overlap. Some flexibility in the timeframes of the samples may also occur. Nonetheless, some coordination is required, including

- Developing an internal communication system to reduce confusion within the utility and with the customers
- Identifying customers appearing in more than one sample
- Developing a protocol for customers who will be contacted more than once
- Developing combined systems for sampling, monitoring management, and quality control to take advantage of economies of scale
- Designing evaluation databases to simplify data comparisons and analyses

The use of models to explain the energy savings may require data leveraging. In this situation, facility inspections or survey data are complemented with billing data, whole-building metering data, or end-use data. These data are then used to develop an explanatory regression model. Examples include billing analysis, conditional demand analysis, and HSEM (Schön and Hamilton 1990).

To perform these types of analysis, various data sets must be linked. This linking is a more stringent requirement than for triangulation methods, where measurement and survey samples may be independent. Linked data requires protocols that ensure a match among facility, end uses, and metered data. This matching can be difficult if some end-uses are included in a master meter and some in a customer's meter, the boundaries for floorspace and energy systems do not match (for example, a store within a mall), multiple meters are used for a facility, or data for multiple fuels are required.

When floorspace and meters do not match, the correction procedure varies depending on whether it is the customer's business boundary, the meter boundary, or the building boundary that is of interest. Which boundary to use is an issue for the analysis method and the method chosen to extrapolate the results. Depending on the choice, the data-collection process may require adding meters or surveying space to create a match. This additional effort increases the costs or time to investigate the match, entails additional data collection, and requires expanded quality control. The best assurance of a reasonable match is to estimate the building energy use from the survey data provided. If the estimated use and the measured use are not within a reasonable tolerance, the survey can be set aside for investigation.

Evaluators may perform a succession of analyses and commit more resources to improve sample designs while minimizing overall costs of an evaluation. For example, an evaluator might do a billing analysis to identify residences that have and use AC; a facility inspection to identify residences with central air conditioners and room air conditioners and to identify the structural characteristics of the residence; an energy audit, a survey, and whole-house monitoring for a large sample of the customers with central AC; and ultimately end-use metering on a selected subsample of the whole-house-metered group. The sample design may be improved at each step.

A good data-management plan is required to support the ongoing data-collection effort as well as the interim analytic steps. Near real-time input and quality control of data are required as well as a database structure that can easily be prepared for analysis while the surveys are still being conducted.

SPECIFYING THE INFORMATION TO BE COLLECTED

Specifying Facility Energy-Use Data

End-use metering requires the specification of the circuits within a facility that are to be metered. The end use must be carefully defined. For example, ventilation is difficult to separate from other elements of heating and AC. The pumps and motors associated with distribution of heated or chilled water may be defined as a heating, cooling, or miscellaneous end use.

End uses rarely are perfectly aligned with electrical circuits. Lighting circuits often include outlets with nonlighting loads. In this case, the evaluator may collect data to account for the plug loads, sample lighting-only circuits, or isolate lighting by rewiring. Obviously, the last method is the most expensive and least practical.

Not all end uses of a single kind in a building may be treated. For example, electronic ballasts may not have been installed in every lighting fixture. It may be difficult to identify and isolate fixtures where the ballasts were installed. The evaluator may have to measure treated as well as untreated lighting. While this captures the savings, it also introduces measurement error.

Measures may have secondary impacts. For example, a reduction in the en lighting reduces cooling requirements in the summer and increases heating req the winter. Both the primary and secondary effects of a measure must be con the lighting example, a well designed data-collection strategy includes measu determine HVAC effects. This measurement may be difficult if both lighting a measures are installed in a facility with a single meter because a single meter measure the combined energy use.

Whole-building-metering procedures are well established in most utilities. When building metering is used for evaluation, however, the ambiguities associated with ac customers, and buildings need to be addressed in the protocol for meter plac Congruence is required among the metered space, the end use, and the area for survey data are collected.

Billing analysis has a number of data requirements, especially when billing data are matc with weather data. Any useful customer-segmentation information on the billing rec (such as SIC code, rate class, or owner-renter designation) should be retained. If multi years of billing data are available, then it is preferable to collect and use all of these da to identify trends. Estimated data in the billing record should be identified as such becau it does not make sense to do regression analysis on data that have been constructed. Meter reading dates are especially important for weather normalization. If weather is not a factor, recording data by month is sufficient. The measured peak kW should be gathered if it is different from billed demand.

Data Specification for Facility Inspections and Customer Surveys

The method of analysis directly affects survey data requirements (Table 10). A bill analysis may not require survey data, while end-use metering may need a prior determination of electrical-system layout. Engineering models are driven by data on physical equipment; such data are obtained through facility surveys. Inputs to statistical models, such as market segment, demographic, or economic data, are usually customer reported. Mail or telephone surveys may suffice, although accuracy and response rate may be increased by an in-person interview. A hybrid model needs physical data and interview information.

Specifying Data for Transferability

Generalizability does not just happen. Data collection must be designed to support the transfer of results to other programs and other utilities where the characteristics of the program and the target audiences may be different; see the discussion of external validity in Chapter 2.

Opportunities for transfer are enhanced when care is taken to ensure that there is a statistically adequate representation of respondents or facilities by space, time, market segments, weather, and economic variables. In selecting segments and the sizes for experimental programs, the population mix of the regions to which the results are to be

DESIGNING THE SAMPLE

Sampling is often the least understood and most poorly executed part of an evaluation. The basic steps in developing a sample design are as follows.

- The population of customers to be represented by the sample, the sample frame, must be defined. Examples of typical frames are all residential customers, small commercial customers with demands less than 40 kW, all residential customers with central AC, or all C&I customers with more than 15 kW of connected fluorescent-lighting load.

- If necessary, the sampling frame is segmented into homogeneous groups on the basis of some characteristic(s), such as kW, kWh, and/or SIC code. SIC code is often used, but many utilities have found it unhelpful because of the wide variation in the size of firms represented within an SIC code.

- The size of the impact variable (e.g., ΔkWh or ΔkW) is estimated for the sample frame as a whole or for the segments. An estimate of savings may be based on engineering estimates or savings reported from other localities.

- Finally, for the population or for the segments, the variance of the impact variable (i.e., savings) must be estimated as well as the variance associated with one's ability to measure the variable. This measurement is usually accomplished by examining a customer database that can be used as a proxy for measured savings (Exhibit 5). Commonly, these proxies are the estimated savings from applications or audits or the variance in customer kWh or kW. Measurement variance and bias are often discounted as unmeasurable or insignificant or as a nonquantified limitation to the sample design. This assumption may not be a good one.

The population size (which is known), the estimated average impact, and the variance of the impact (which is estimated) provide the basis for the calculation of sample size requirements to meet the desired sampling error, specified as a precision (such as ±10%) with a specified level of confidence (such as 90%).

Trading Off Sources of Error

Sources of error are manifold. If precise measurement of savings were possible, then sampling (as compared with conducting a census) would be the only cause of error. However, measurements are also subject to error. Much of the data collected in an evaluation are treated as observed data when in fact they contain measurement error. For example, operating hours for a commercial building are observed during an onsite survey but may be reported for off-peak periods. These off-peak data are treated as measurements but contain an unknown amount of error. Reported evening operating hours in commercial buildings are often lower than true levels.

Exhibit 5. Calculating sample sizes for an evaluation

Utility A plans to run a weatherization program that is similar to that of Utility B. Utility B had the following results when it conducted its weatherization program.

Table 11. Mean consumption or savings (kWh/year) and standard deviation (SD)

	Mean consumption or savings (kWh/year)	Standard deviation
Preprogram electricity use	26,500	8,480
Postprogram electricity use	21,480	8,130
Savings	5,020	4,270

Utility A wants to know how large a sample it needs to determine the savings for its program. Because the programs are nearly identical, Utility A assumes that the savings from its program will be similar to the savings from Utility B's program. It therefore decides that it can use the information from Utility B's program to determine the sample size for its program based on the following formula:

$$n = (t^2s^2N)/(E^2N + t^2s^2) \ ,$$

where n = sample size; t = the t value associated with a sample size and level of confidence taken from a t-table in a statistics textbook; s = estimated SD; N = the population size; and E = required level of accuracy.

Initially, the evaluator assumes that the sample size should be based on the postprogram consumption (mean = 21,480 kWh). She then realizes that the postprogram consumption was not of interest but rather the savings and the SD for the savings. She also realizes that management will ask questions about the sample size, so she constructs the following table to help management determine how to best use evaluation resources.

She assumes three levels of confidence (0.1, 0.05, and 0.01), three levels of accuracy (10%, 5%, 1%), and three levels of population (10,000, 20,000, 100,000). The results are displayed in the following table that shows sample sizes as a function of confidence level and accuracy:

Exhibit 5. (Continued)

Table 12. Sample sizes as a function of confidence level and accuracy

Confidence	Accuracy 10%	Accuracy 5%	Accuracy 1%
	Population = 10,000		
0.1	117	453	5,349
0.05	192	725	6,544
0.01	376	1,352	7,911
	Population = 20,000		
0.1	118	464	
0.05	194	753	
0.01	384	1,450	
	Population = 100,000		
0.1	119	472	
0.05	195	775	
0.01	389	1,539	

The computation for the first cell is:

$$n = 1.282^2 4270^2 10000/(502^2 10000 + 1.282^2 4270^2)$$

Several things are clear from this table.

- The size of the population has little effect on sample size.
- The required sample size for 1% accuracy is so large as to be impractical to implement.
- The trade-offs between accuracy and confidence level are about equal.

Other errors occur during data handling. Data are often mistakenly entered or misread. While good quality-control procedures can reduce data errors, they will not eliminate them.

Measurement accuracy and sample size can be traded off to minimize sources of error within a fixed evaluation budget. For example, end-use metering is accurate but expensive, and samples are usually kept small to reduce costs. The consequence is a large sampling error. Within the same budget, sampling error might be reduced by using a large sample in conjunction with load meters, field inspections, and hybrid statistical engineering models that have a lower per unit cost. Progressively larger samples can be afforded when field inspection is used with engineering estimation, mail surveys are used with statistical modeling, and bill analysis is used with a census. These larger samples have progressively lower sampling errors but progressively larger unit measurement errors.

Defining Precision Requirements Sensibly

Defining precision requirements is a critical element in determining sample sizes. In making this determination, one must consider the importance of the measurement relative to evaluation objectives. For example, an estimate of lighting electricity savings may require measurement of peak kW savings and operating hours. The peak impact of a high-efficiency lamp can be determined with a fair amount of certainty. The midday operating pattern may be reasonably predictable, but the evening operating pattern may be much more variable. For a utility with an evening peak, the precise measurement of evening operating hours would be more important than it would be for a utility with an afternoon peak. A larger sample size may be required for the evening-peaking utility.

In a recent study, building data were collected for a load-research sample. As part of the sample-planning process, statistical, engineering, and hybrid models were examined for accuracy and cost in relation to metering options. The study also assessed the relative importance of individual measurements to the impact variables. Survey data from a large sample were determined to be lowest in cost but were found to have a relatively large uncertainty. Hybrid approaches applied to a subsample of critical building types and measures substantially reduced the potential bias. Where uncertainty and importance remained, end-use metering appeared to be the best option (XENERGY 1990).

ENSURING DATA QUALITY

Support, Training, and Management of Survey Takers

When possible, experienced staff should be used for field and telephone surveys. When this is not possible, evaluators should recruit people who have a basic understanding of energy and who know how to inventory and evaluate building envelope and energy systems. In addition, survey takers should be familiar with the programs, processes, and applications referred to in the survey. Equally important are people with social skills.

Support materials for survey takers should include, but not be limited to, a program description and introduction, line-by-line instructions for the survey instrument, technical support materials on equipment and devices, a protocol for dealing with issues that may arise in the field with particular attention to sampling, a set of procedures for scheduling conflicts, and account-contact guidelines.

Training survey takers is one of the most important aspects of a quality-control program. Even if the survey takers have extensive energy-survey experience, a training session is essential to instruct them on the specifics of the particular project and to brief them on the issues that they may encounter in the field. The training session, in conjunction with the survey training manual, helps to ensure that each survey taker interprets the questions in the same fashion, thereby providing the required consistency. Survey takers should practice on each other and perhaps on a group of volunteers before going into the field.

A supervisor should observe survey takers from time to time to be sure that they continue to ask questions appropriately and follow procedures. Feedback about survey-taker performance should also be solicited from a randomly selected group of those being visited in the field. Questionnaires should be edited immediately after they are returned, and any patterns of omission or error should be called to the survey taker's attention. Finally, supervisors should review information on the face sheets of surveys (e.g., the amount of time spent conducting the interview) to see if these data give indications of interviews being too short or too long or the timing of interviews being inconsistent with estimated travel requirements.

Data Quality Control

Despite training, errors will occur in reporting and in data entry. Several methods of prevention are available. Double entry of data minimizes input errors. Range checks can catch reporting and input errors. More sophisticated validity checks, which compare data with a set of predetermined parameters (such as watts per square foot, occupants per square foot, watts per fixture, etc.) can be programmed into software. Finally, consistency checks can verify predetermined relationships between survey questions. For example, a centrifugal chiller system in a 10,000-ft^2 building is unusual, suggesting the need for further investigation.

Survey data errors can be identified and reduced by offering a computer-generated energy analysis to the customer. This is a particularly useful practice in mail surveys. This inducement encourages the customer to exercise care and provides an additional error check by virtue of the customer's review of computed results based on the data he/she provided. In addition to reduced error rates, this offer may increase residential mail survey response rates as much as 45%, reducing self-selection bias.

Quality Control of Experimental Procedures

Logs should be maintained for all customers being recruited, and the status of all customers participating in a survey. Tracking the status serves as a record to ensure a survey's statistical value, as a management tool to evaluate the effectiveness of survey resources, and as an aid to the utility in dealing with customer issues.

The data-collection manager should conduct ongoing analytic activities to identify potential problems with a survey. Quality-control procedures include testing variance assumptions, correcting survey problems, and adjusting analytic methods.

The research plan is designed to measure impacts to a desired precision and confidence level. The estimate of sample size required to meet this level includes variance assumptions for the sample population. Because of the uncertainties about the actual variance, sample sizes may be too high or too low. An ongoing analysis of the variance of key variables during data collection might identify the need to adjust sample size. The variance data can

ANALYZING DATA

Daniel Violette

INTRODUCTION

This chapter discusses data-analysis methods for estimating the impact of DSM programs on kW or kWh consumption. For any given program, several candidate analysis techniques will exist, each with its own strengths and weaknesses. The selection of an appropriate method will depend upon:

- Specific program design
- Available prior information on program impacts
- Program's priority in the utility's DSM portfolio
- Implementation stage (whether it is a pilot, new, or mature program)
- Objectives of the evaluation
- The technologies and customer segments targeted by the program

Evaluations can be based predominately on engineering methods, billing-data analyses, metered data, or combinations of these methods. While most data analyses can be classified as falling into one of these categories, the use of multiple methods is a trend. These approaches go beyond conducting parallel engineering and billing-data analyses. Instead, they often mean leveraging and integrating data across more than one approach. This technique could involve using estimates from an engineering study as inputs to statistical models, or it could involve using relationships derived from a statistical analysis of consumption data within an engineering model.

An impact evaluation has two purposes, to measure the *change* in participant electricity use and to identify the factors that explain the observed change, especially the effect of the utility DSM program. The challenge for evaluators is to determine how this intervention in the market for energy services influenced the energy use of participants and, often, the energy use of nonparticipating customers as well.

Because the purpose of an impact evaluation is to measure changes in energy use, a baseline must be identified from which a change can be measured or estimated. Determining this baseline is a critical step. Monthly billing data (or meter data) can measure *current* energy consumption exactly; however, knowing the postparticipation consumption does not provide an estimate of the change in energy use caused by the program. To determine impacts, an estimate of the amount of the energy that would have been used in the absence of the program is needed; see Fig. 6 in Chapter 2. This is accomplished by establishing a baseline to serve as a proxy for the consumption that would have occurred without the DSM program.

Many of the estimation problems associated with impact evaluation are related to the selection of an appropriate baseline. Three such problems (self-selection, free riders, and free drivers) can be viewed as biases in the baseline (see Chapter 8 of this volume; EPRI 1991b; Violette, Ozog, and Wear 1991). Self-selection bias occurs when program participation is voluntary. In this circumstance, systematic differences may occur between nonparticipants, who either chose not to participate or were unaware of the program, and participants. Free riders and free drivers are important when the actions undertaken by the comparison group may not accurately reflect the actions that would have been undertaken by participants without the program.

STATISTICAL COMPARISON APPROACHES

This section outlines several simple-comparison approaches, shown in Table 7 of Chapter 2, plus methods that use other data available at the utility.

Three Comparison Options

The time-series approach compares the participants' pre- and postparticipation energy usages. The energy use of the participants prior to their participation in the program is used as the baseline energy use. Energy savings is estimated by:

$$\text{Savings} = Q_b - Q_a \ ,$$

where Q_b is the quantity of energy used before participating in the program and Q_a is the quantity used after participating. The before and after time periods may be annual periods or heating and cooling seasons.

In this case, the comparison group is comprised of the same set of participating customers. Among the comparison approaches, this method has the advantage of using a comparison group that is nearly identical in its dwelling and demographic characteristics because they would not be expected to change substantially over a 2- to 3-year period. However, this approach cannot capture other factors that may change with time, in particular electricity prices or weather. In addition, this comparison approach, by not using the change in energy use for nonparticipants, is not able to estimate natural conservation.

The cross-sectional comparison of participants with nonparticipants examines postprogram energy use across two customer groups:

$$\text{Savings} = Q_{np} - Q_p \ ,$$

where Q_{np} is the quantity of energy consumed by nonparticipants and Q_p is the quantity consumed by participants. In this case, the baseline is the nonparticipant energy use. This approach assumes that, in aggregate, the nonparticipant comparison group is identical to the participant group in all respects except for program participation. The advantage of this approach over the time-series method is that, because one time period is used, no biases result from factors that changed over time (e.g., weather conditions or prices). However, a representative comparison group may be difficult to obtain.

Combined time-series/cross-sectional approaches combine the advantages of both approaches and, most importantly, also control for natural conservation. These approaches compare *changes* in energy use over two time periods for two groups:

$$\text{Savings} = [Q_{bp} - Q_{ap}] - [Q_{bnp} - Q_{anp}] \ ,$$

where Q_{bp} and Q_{ap} are the quantities of energy consumed by participants before and after participating in the program, respectively. Q_{bnp} and Q_{anp} are the quantities of energy consumed by nonparticipants for the same time periods.

A comparison approach that uses percentage changes in energy use is often used:

$$\text{Percent savings} = [(Q_{bp} - Q_{ap})/Q_{bp}] - [(Q_{bnp} - Q_{anp})/Q_{bnp}] \ .$$

This percentage approach is generally preferred to the absolute-change approach because percentages help to normalize the estimated natural change in consumption for differences in the average size of customers between the participant and comparison groups.

The time-series/cross-sectional approaches are the best of the comparison approaches. Correcting the program savings estimates by the savings that occurred among nonparticipants helps to control for factors that change over time and influence energy use. Also, to the extent that the propensity for natural conservation is the same for participants and nonparticipants, the approach also controls for natural conservation (Rogers 1989; BPA 1990).

The drawback to this approach is that it assumes that factors such as weather and energy prices affect both participants and nonparticipants equally. This may not be the case. However, even the most sophisticated multivariate, simultaneous-equations approaches discussed later in this chapter are nothing more than variants of this simple-comparison concept. The sophisticated approaches simply attempt to develop better proxies for baseline energy use.

Comparison Approaches Augmented by Other In-House Data

All the comparison approaches discussed above are limited in their ability to control for those factors (aside from participation) that affect energy consumption, such as weather, electricity prices, appliance stocks, and dwelling characteristics. However, utilities generally have information on weather and electricity prices in-house, so it is possible to control for these factors without undertaking additional customer surveys. A simple regression framework can accomplish this control where program participation, weather variables, and energy prices are used as explanatory variables. One widely used approach to control for weather is PRISM (Fels 1986; Fels and Reynolds 1991; Brown et al. 1991).

Besides weather, other confounding variables can be incorporated into the comparison techniques without having to rely on customer surveys. One such variable is electricity price. If the evaluation spans a 2- to 3-year period, the impact of changes in electricity price on consumption can be important. Because most utilities have this information within the billing file, obtaining it requires minimal effort. Other potential variables include house type (for residential programs), SIC codes (for commercial programs), or other customer and premise data contained in the utility's master customer-account files.

Statistical Precision

Designs and sample sizes that estimate *levels* of consumption with a high degree of precision must be distinguished from studies designed to estimate DSM savings, (i.e., a *change* in energy use). The large differences in precision between estimates of levels and estimates of changes can be illustrated with a simple example.

Assume a DSM program in which end-use metering is conducted on 100 participants and on 100 nonparticipants. The average consumption for participants was 25 kWh per day with a SD of 8 kWh per day. (Estimates of kWh savings per day are often used because they eliminate the need to control for different numbers of days in billing cycles, months, or heating seasons.) The mean use for nonparticipants was 30 kWh per day with a SD of 10 kWh per day. The savings estimate is the difference between these two mean consumption figures: 5 kWh per day or 16.7% of the mean nonparticipants' consumption.

From these data, the precision with which the mean consumption levels are estimated can be calculated for both groups. The following formula is used to calculate the absolute error of the estimates of means:

$$\text{Absolute Error} = t_{cr} \sqrt{\frac{s^2}{n}} \quad , \qquad (4\text{-}1)$$

where t_{cr} is the critical t-value associated with the confidence level (in this case, 1.98 for a 95% confidence level), s is the SD of the estimate (8 and 10 kWh, respectively), and n is the sample size (in this case, $n = 100$).

The end-use metering resulted in an absolute error of 1.98 for the estimated mean consumption for nonparticipants and 1.6 for participants. The relative precision is defined as:

$$\frac{\pm \textit{Absolute Error}}{\textit{Estimated Mean}} \ .$$

The relative precision is ±6.5% for both participants and nonparticipants. Given these very precise estimates of mean consumption levels, the next issue is to determine the accuracy of the savings estimate based on the difference between these means. This accuracy is calculated by determining the absolute error of the savings estimate (i.e., the 5-kWh difference between the two means):

$$\text{Absolute Error}_{diff} = t_{cr} \sqrt{\frac{s_p^2}{n_p} + \frac{s_{np}^2}{n_{np}}} \ ,$$

where s_p and s_{np} are the SDs of participants and nonparticipants, respectively, and n_p and n_{np} are the sample size of participants and nonparticipants, respectively.

A 95% confidence level gives an absolute error of 2.5 kWh. The relative precision is found by dividing this absolute error by the 5-kWh/day savings, giving a ±50% relative precision. This large absolute error for the estimated savings contrasts sharply with the ±6.5% precision around the mean-consumption estimate for each group. The example illustrates that research designed to produce precise estimates of levels of consumption may not produce the hoped-for precision in DSM-impact estimates (i.e., the change in consumption).

USING PAIRED DATA TO ENHANCE PRECISION

Paired data occur when the sample observations contain two or more data elements. Examples of paired data can include (1) a prior estimate of savings and the floor area of the same building taken as a pair or (2) an engineering estimate of energy savings paired with a field estimate of energy savings based on billing data. In both instances, two observations for the same customer form a pair. Such paired observations can, in certain circumstances, provide estimates with smaller variances (Walpole and Myers 1985).

Combining Population and Sample Information

Certain data that might be available for the population of participants can be combined with sample data to increase the precision of estimated savings. The procedure illustrated here consists of normalizing energy use over a *known* variable to reduce variance in the estimates. One potential cause of variation in energy savings across buildings may be the size of the building. Large savings are more likely in large buildings. If this cause of variation can be eliminated, then the variance around the estimated savings can be reduced. For example, kWh savings per square foot may have a smaller variance and, therefore, can be estimated with greater precision. However, this procedure requires estimates of floor area for every customer in the sample, and if estimates of total program savings are to be made, these data are needed for all participants. Because energy audits are conducted as part of many DSM programs, such information is often available for all participants.

To illustrate, consider a program that has 250 participants, with meters installed to measure before and after electricity consumption for five of them. Two cases are presented: Case 1 (Column 1 of Table 13) uses sample data only and Case 2 (all three columns of Table 13) uses information from the energy audits.

Table 13. **Estimates of electricity savings without and with additional information on building floor area**

	Estimated savings, kWh (metered data) (1)	Floor area (ft^2) (2)	Estimated savings (kWh/hundred ft^2) (3)
Customer 1	25,000	15,000	166.7
Customer 2	45,000	25,000	180.0
Customer 3	20,000	12,000	166.7
Customer 4	50,000	31,000	161.3
Customer 5	15,000	9,500	157.9
Mean	31,000	18,500	166.5

With only the estimates in Column 1, the mean of the estimated savings for this sample is 31,000 kWh with a standard deviation of 15,600 kWh. A 90% confidence interval (CI) around this estimate is ±2.13 times the SD divided by the square root of the sample size, or 31,000 ± 14,860 kWh. Thus, the level of precision for this estimate is ±48%.

The second and third columns of Table 13 illustrate the use of additional information available on the population of program participants from audit records, in this case the floor area of each building. Now it may be possible to estimate program energy savings more precisely. This example shows that some of the variation in energy savings is explained by the size of the building. The mean savings per 100 square feet is 166.5 kWh with a SD of 7.5. The precision around this estimate is ±4.3%, significantly less than the ±48% from the unnormalized sample.[2]

This example illustrates the advantage that can be gained by using information available on the population of participants to help explain the variation in savings. Often, data like these are available from audits conducted prior to program participation, or they can be collected as part of program implementation. Because program savings were, in this example, correlated with the building square footage, the gain in precision was significant.

---

[2]This example does not necessarily assume that the floor area for each building is measured precisely. Measurement error in these estimates will contribute to a larger SD in the estimated savings per square foot. As long as the floor-area data are positively correlated with energy use, the precision of estimated savings will be increased.

Difference and Ratio Estimates

Difference and ratio estimates can increase the precision of estimates of energy savings from conservation measures. These techniques have been applied in many areas of statistical analysis (Arkin 1974). In fact, one approach used in some evaluations, Statistically Adjusted Engineering (SAE) estimates, is a variant of the ratio estimator.

Difference- and ratio-estimation approaches can be viewed as procedures for "auditing" the accuracy of the prior engineering estimates. This auditing is accomplished by using field measurements of a sample from this population. Because many DSM programs, particularly C&I programs, develop detailed engineering estimates as part of program implementation, difference and ratio estimators have the potential for widespread use in program evaluation.

The difference estimator *modifies* the prior population engineering estimates by calculating an estimated mean difference between field and prior-engineering estimates using a sample of customers. For example, assume a commercial DSM program with 750 participants. The engineering estimates made as part of the program found that the average savings per participant for the total population would be 34,000 kWh. The program metered a sample of five participants to obtain field estimates of savings (Table 14).

Table 14. **Electricity savings (kWh) from meter data and engineering estimates**

	Estimated savings (meter data)	Estimated savings (engineering)	Difference (meter-engineering)
Customer 1	25,000	29,000	-4,000
Customer 2	45,000	44,000	+1,000
Customer 3	20,000	26,000	-6,000
Customer 4	50,000	48,000	+2,000
Customer 5	15,000	13,000	+2,000
Total	155,000	160,000	-5,000
Mean	31,000	32,000	-1,000
SD	13,928	--	3,742

Based on these data, the average engineering estimate is 1,000 kWh more than the metered estimate. (Note that the mean engineering estimate for the population is 34,000 kWh, but the mean for the 5 customers in the sample is only 32,000 kWh.) Therefore, an adjustment factor based on this estimated difference is applied to the mean engineering estimate for the population. With an estimated difference of -1,000 kWh, the resulting estimate of the population mean is 34,000 + (-1,000) = 33,000 kWh.

If the conventional estimation approach is employed, with only the mean value for the metered sample of customers, the resulting estimate of the sample mean is 31,000 kWh. The precision of this estimate is found by:

$$SD_{conv} = \sqrt{\left[\frac{\Sigma(X_i - Mean_x)^2}{(n - 1)}\right]} .$$

For the data in the example, the SD for the conventional approach is 13,928 kWh. With a 90% confidence level, the relative precision is 43%.

To measure the sampling precision of the difference estimate, the SD of the difference between the metered estimates and the engineering estimates is determined by:

$$SD_{diff} = \sqrt{\left[\frac{\Sigma(X_i - (Y_i + Diff))^2}{(n - 1)}\right]} ,$$

where X_i is the metered estimate for observation i; Y_i is the engineering estimate for observation i; Diff is the difference adjustment factor; and n is the sample size. With the data in Table 14, the resulting estimated SD for the difference estimator is 3,742 kWh. The precision, with a 90% CI, is ±11%.

For these hypothetical data, incorporating prior engineering estimates to create a difference estimate produces a program energy-savings estimate that is nearly four times more precise than the conventional approach using only the meter data for the sample.

While the difference estimate uses the difference between the meter estimate and the engineering estimate, the ratio estimate uses the ratio of the metered estimate to the engineering estimate. The applications of these two methods is similar; the choice between the difference and ratio approaches depends on the relationship between the metered estimates and the engineering estimates (EPRI 1991a; Wright and Townsley 1990). If the size of the difference between the two is independent of the size of the savings (i.e., a large engineering estimate does not imply a large difference between the engineering estimate and the metered estimate), then the difference estimator is more precise. In the extreme, if the engineering estimate is always a fixed magnitude different from the metered estimate, then the SD of the difference estimate is zero.

If the size of the difference between the engineering and metered estimate is related to the size of the engineering estimate, then the ratio estimate is generally more precise. If the engineering estimate is consistently different by a certain percentage (e.g., it always overstates savings by 20%), then the SD of the ratio estimate is zero. In practice, if a difference estimate can be calculated, then a ratio estimate can also be calculated. The analyst can then calculate both difference and ratio estimates and select the approach that produces the smallest standard error of the estimate.

In the above example, the predicted energy savings were not normalized for floor area. The normalization method presented above can be combined with statistical audit procedures to further increase the precision of estimates. In this case, both the engineering estimates and the meter data would have to be expressed in terms of savings per square foot.

SAE Estimates as Ratio Estimates

SAE estimates are a type of ratio estimate for which the paired data encompass a field-derived statistical estimate and a prior engineering estimate of savings. SAE applications fall into two categories: (1) auditing the accuracy of engineering-based program-saving estimates and (2) disaggregating whole-building-metered loads into individual end-use loads (EPRI 1991a). The example presented here focuses on DSM program savings.

SAE estimates are produced by using engineering estimates as normalizing variables or, when more than one DSM measure is being considered, as independent variables in a regression equation. The basic non-SAE model is:

$$\Delta Energy_t = Energy_t - Energy_{t-1} = \beta_1(Part_1) + \beta_2(Part_2)$$

or

$$Energy_t = \beta_1(Part_1) + \beta_2(Part_2) + \lambda(Energy_{t-1}) \ ,$$

where λ is the coefficient on the lagged energy variable, $Energy_{t-1}$, and $Part_1$ and $Part_2$ are (0,1) binary variables that indicate participation in more than one DSM program (i.e., programs 1 and 2), where the variable takes on the value of 1 if that customer is a participant and is zero otherwise. Alternatively, $Part_1$ and $Part_2$ could refer to different DSM measures.

Under the SAE approach, the basic model is:

$$Energy_t = \beta_1(Eng_1) + \beta_2(Eng_2) + \lambda(Energy_{t-1}) \ .$$

Now the participation variables are engineering estimates (i.e., Eng_1 and Eng_2) of the savings from different DSM measures or savings associated with participation in programs 1 and 2, respectively. In either case, program-specific or measure-specific engineering estimates of savings must be available. SAE estimates can also be applied in a multivariate format. The only difference is the addition of other explanatory variables, X_i, to the above equation.

The interpretation of the regression coefficients in the SAE model is different from the interpretation of those in the first model. A coefficient of 0.5 indicates that, on average, only 50% of the predicted engineering estimates were in fact realized, based on the measured consumption data.

The use of engineering estimates of savings as independent variables in the regression equation instead of dichotomous 0-1 variables has several potential advantages. First, it is a way to compare statistically-derived estimates with engineering estimates. Second, the direct incorporation of the engineering estimates into the equation brings more information to bear on the estimation problem. If one building, according to the energy audit, has greater potential savings than does another building, then this information is reflected in the variable representing participation (i.e., the engineering-savings estimate). If this information is hierarchically accurate (that is, the building does have a higher savings potential and does achieve higher savings), then the precision of the estimated savings should increase. If the engineering model is so inaccurate that it is unable even to hierarchically rank buildings by savings potential, then the SAE model estimates may be less precise than simply using (0,1) participation variables. This condition is not very restrictive, and SAE estimates will likely increase the precision of savings estimates in most instances. However, this thesis can be tested statistically. Simply estimating the equation with (0,1) indicator variables and also estimating the model with the engineering estimates will show which specification produces the more precise estimate. If the engineering estimates are accurate enough to hierarchically rank customers by savings, then the SAE formulation will produce savings estimates with smaller SDs and higher *t*-values than will the non-SAE model formulation.

MULTIVARIATE-SAVINGS-MODEL SPECIFICATIONS

The comparison approaches discussed above require data that most utilities typically have and produce useful information about the impacts of a DSM program. However, with these approaches, the researcher cannot know whether the participant and nonparticipant groups are truly comparable. The only data available to confirm the similarity of the two groups are billing data and other in-house information. Adding customer-specific information (e.g., from surveys) may increase the confidence in impact estimates. In some cases, such additional information is obtained from energy audits of samples of participants and nonparticipants (CMP 1990; BPA 1990). To date, most multivariate models have been developed for residential programs; however, applications to commercial-sector programs are increasing (Consolidated Edison 1991; Central Hudson Gas and Electric 1991; San Diego Gas and Electric 1991; Train and Ignelzi 1987).

The availability of additional customer-specific data allows for the use of multivariate models that can incorporate more structure in terms of engineering relationships, building characteristics, and customer behavior and attitudes. This additional information can be important when program savings are expected to be small compared to the customer's total energy consumption. A model to isolate program impacts on the order of 5% of total consumption requires that all available information be used, including a well-defined structure and as much information from engineering principles as possible.

Evaluators can estimate total energy-use models (cross-sectional models) and change models (pooled time-series/cross-sectional models). One total-energy-use model and two forms of change models are presented in Exhibit 6.

Exhibit 6. Specifications of energy-use models

Three types of energy-use models can be used for DSM-impact evaluation, as illustrated here. Assume that there are three end uses [space heating (SH), AC, and water heating (WH)], that program participation is denoted by a 0,1 indicator variable (Part), and that participation occurs at the end of 1990.

1. **Cross-Sectional Model**

$$E_{1991} = \beta_0 + \beta_1 SH_{1991} + \beta_2 AC_{1991} + \beta_3 WH_{1991} + \beta_4 Part + \epsilon_{1991}$$

2. **Constrained-Change Model**

$$E_{1991} - E_{1990} = [\beta_0 - \beta_0] + [\beta_1 SH_{1991} - \beta_1 SH_{1990}] + [\beta_2 AC_{1991} - \beta_2 AC_{1990}] + [\beta_3 WH_{1991} - \beta_3 WH_{1990}] + \beta_4 Part + [\epsilon_{1991} - \epsilon_{1990}]$$

or

$$\Delta E = \beta_1(\Delta SH) + \beta_2(\Delta AC) + \beta_3(\Delta WH) + \beta_4 Part + \Delta\epsilon$$

3. **Flexible-Form Model**

$$E_{1991} = [\beta_0 - \lambda_0\beta_0] + [\beta_1 SH_{1991} - \beta_1\lambda_1 SH_{1990}] + [\beta_2 AC_{1991} - \beta_2\lambda_2 AC_{1990}] + [\beta_3 WH_{1991} - \beta_3\lambda_3 WH_{1990}] + \beta_4 Part + \lambda_4 E_{1990} + [\epsilon_{1991} - \lambda_5\epsilon_{1990}]$$

Here, λ_0, ..., λ_5 are adjustment factors. In all cases, the effect of the program on energy use is given by the coefficient on the participation variable (β_4).

Total Energy-Use Models

Preparticipation data simply are not available in some cases, such as for new-construction programs. For these programs, the only option is to use a cross-sectional framework in which the postparticipation energy use of a participant group is compared with the energy consumption of a nonparticipating group. This comparison can be done with a multivariate-regression model in which total energy use is modeled and a participation variable is used as an explanatory variable.

The total-energy-use model should include all measurable factors that influence electricity consumption. Any omitted factor that varies systematically across participants and nonparticipants will bias the estimates of program impacts. Although different specifications of models can be used, one approach that is theoretically sound involves use of a conditional-demand specification. Total electricity demand is modeled as a function of the electricity used in each end use.

A model of residential energy use illustrates the approach. The household's total electricity demand at time t (E_t) is the sum of the electricity used for space heating (SH_t), air conditioning (AC_t), and water heating (WH_t) and an error term (ϵ_t):

$$E_t = SH_t + AC_t + WH_t + \epsilon_t \quad . \tag{4-2}$$

In practice, electricity use for each end use is itself a function of variables that represents customer behavior and engineering principles (EPRI 1991a; Parti and Parti 1980). Applications of the conditional-demand model for estimating DSM-program savings with a total energy model can be found in Wisconsin PSC (1989), CMP (1990), and Violette and Ozog (1989).

Change Models: Pooled Cross-Sectional/Time-Series Models

The total-consumption models discussed above are cross-sectional (i.e., they analyze total energy consumption at one point in time). Models that include changes in energy use over time are likely to be more appropriate for DSM-program evaluations. Pooled cross-sectional/time-series models have the following form:

$$\Delta E = \beta_0 + \beta_1 \Delta SH + \beta_2 \Delta AC + \beta_3 \Delta WH + \epsilon \quad , \tag{4-2a}$$

where Δ denotes the change in the value of that variable (e.g., $\Delta E = E_t - E_{t-1}$).

For each end-use variable, a change can result from changes in one or more of the individual components of that variable. For example, the energy used for SH can change because of changes in the heating equipment, building insulation, occupant behavior, or weather. In some cases, no changes will occur in the individual variables so the composite change variable will not appear in the model. Equation (4-2a) is referred to as a constrained model because it constrains the coefficient (i.e., the energy-use intensity) for each end use to be constant between the t and t - 1 periods being analyzed.

A more flexible form of the change model uses a lagged dependent variable as an independent variable (i.e., E_{t-1} appears on the right-hand side of the equation) and introduces adjustment factors (λ_i) for each end use.

Table 15. **Electricity-savings, unconstrained-change model**[a]

Independent variable	Coefficient	(*t*-values)
Intercept	1256.42	(4.73)
Participation variable	-545.94	(-4.30)
Number of people at home during the day in 1987	527.65	(4.99)
Number of people at home during the day in 1985	-389.97	(-3.63)
SH[b] times heating degree days, 1987	0.00047	(3.83)
SH[b] times heating degree days, 1985	-0.00057	(-4.81)
SH[b] times average price, 1987	-0.0215	(-12.24)
SH[b] times average price, 1985	0.0237	(12.03)
SH[b] times income, 1987	0.00020	(7.46)
SH[b] times income, 1985	-0.00020	(-7.35)
Company #1 indicator variable	-1234.83	(-5.93)
Company #2 indicator variable	-910.23	(-4.78)
Electricity consumption, 1985	1.03	(77.72)
Number of observations	1169	
R^2	0.90	
Mean of dependent variable	12,460	

[a]The dependent variable is postparticipation electricity consumption (1987). The preparticipation period is 1985, with 1986 being the participation period.

[b]SH is a variable representing the estimated surface area of the home.

Source: New Jersey Conservation Analysis Team (1990).

this may be difficult to accomplish, even with the sophisticated estimation techniques designed to address self-selection.

Two general approaches for addressing self-selection within multivariate models are available. The first approach is based on simultaneous-equations methods; the second uses the properties of the cross-sectional/time-series change models to reduce the likelihood of self-selection bias.

Simultaneous-Equations Approach

Self-selection can be viewed as a situation in which the amount of energy a customer uses is a function of program participation and the decision to participate is a function of energy use. Thus, participating in a DSM program and energy use are joint decisions.

Approaches for addressing self-selection based on a simultaneous-equations framework require the estimation of two equations, a discrete-choice participation model and an energy-use equation such as the regression models discussed above. Among the approaches for explicitly addressing self-selection bias, one easily applied method is the selectivity-correction-factor technique. This technique, developed by Heckman (1978) and refined by Dubin and McFadden (1984), involves the estimation of a participation model and a multivariate-regression model of energy use. In theory, if the researcher can incorporate into the energy-use equation all the variables that characterize participation, then self-selection bias will be minimized. Many of the variables that influence the decision to participate are difficult to observe (e.g., attitudes and beliefs). The selectivity-correction technique uses a participation model to estimate the probability of participation. Then, based on the estimated probability of participation, a selectivity-correction term is developed. The general form of this correction term is:

$$C_i = \left[\frac{\hat{P}_i \ln \hat{P}_i}{(1 - \hat{P}_i)} + \ln \hat{P}_i \right] , \qquad (4\text{-}4)$$

where $\hat{P}_i$ is the estimated probability of participation from the participation model. The new energy-use equation is:

$$\Delta Energy = \beta_1(PART) + \beta_2(C_i) + \beta_3(x_i) . \qquad (4\text{-}5)$$

This is a multivariate regression model, as discussed in the previous section, with one change: C_i is added as an explanatory variable. Calculating the C_i is straightforward:

- Estimate a discrete choice participation model with a sample of participants and nonparticipants, data on customer attitudes, and other factors believed to influence the decision to participate in the program.[3]

- Use the participation model to estimate the probability of participation for each customer in the sample.

- Use the probabilities of participation for *each* customer in the sample to calculate the C_i for each customer.

- Include the calculated C_i variable in the energy-use regression equation.

---

[3]A discrete-choice model uses discrete outcomes as the dependent variable. Here, the dependent variable is 1 if the customer is a participant and is 0 if that customer is a non-participant. The independent variables include factors that influence customer decisions to participate, including income, expected energy savings resulting from participation, and awareness of the program.

The selectivity-correction approach introduces the term C_i into the equation, which allows the researcher to test for bias. Violette and Ozog (1989) pointed out several aspects of this technique that influence the interpretation of the findings. Most importantly, a large, statistically significant coefficient on the selectivity correction term does not necessarily imply that the estimates of program energy savings have a large bias. In some cases, little correlation exists between the magnitude of the coefficient on the selectivity term and the size of bias in the energy-savings estimate. The only way to examine this is to estimate the model with and without the selectivity factor and then compare the resulting savings estimates. Finally, the bias can be either positive or negative (i.e., it can increase or decrease estimated program savings).

Implicit Correction for Self-Selection: Pre/Postregression-Analysis Techniques

A second approach to self-selection uses a change model based on pooled time-series/cross-sectional data. Under a reasonable set of assumptions, the change models presented in the previous section can reduce self-selection bias (Heckman and Robb 1985; EPRI 1991a). This technique exploits the fact that an evaluation is only concerned with how energy use changes over time and not what determines total energy use. By investigating only the factors that alter energy use, one can mitigate the effects of self-selection bias without using sophisticated models.

To understand this approach intuitively, assume that participants in a conservation program are more likely to consume less energy without the program than nonparticipants are. If this propensity to consume less energy is constant over the short run, then a difference model cancels out this propensity to consume less energy, eliminating the bias.

The implication of this result is that by collecting data on at least one period of preprogram energy use, researchers can produce an estimate of DSM impacts that may not be affected by self-selection. However, this approach does not correct for self-selection if participants' and nonparticipants' change in energy consumption exhibits a systematic difference with time. For example, if participants consistently take more conservation actions over time than nonparticipants, then participants not only will have lower levels of energy consumption, but also will have more rapidly changing energy consumption. Figure 8 depicts such a situation. Because the rate of change in energy use over time is different for the two groups, this difference between participants and nonparticipants does not cancel out in a two-period model. Under these circumstances, a pre/postmodel will not fully correct for self-selection bias, and the use of the self-selection correction factor may be warranted.

Two methods for mitigating self-selection bias have been presented. The choice of which approach to use depends on several factors. First, the selectivity correction technique can be used with both cross-sectional and pooled cross-sectional/time-series data, while the change in energy use approach can only be used with pooled cross-sectional/time-series data. Thus, the selectivity-correction technique is particularly useful for new-construction programs. Also, the potential for self-selection bias is much greater in cross-sectional models.

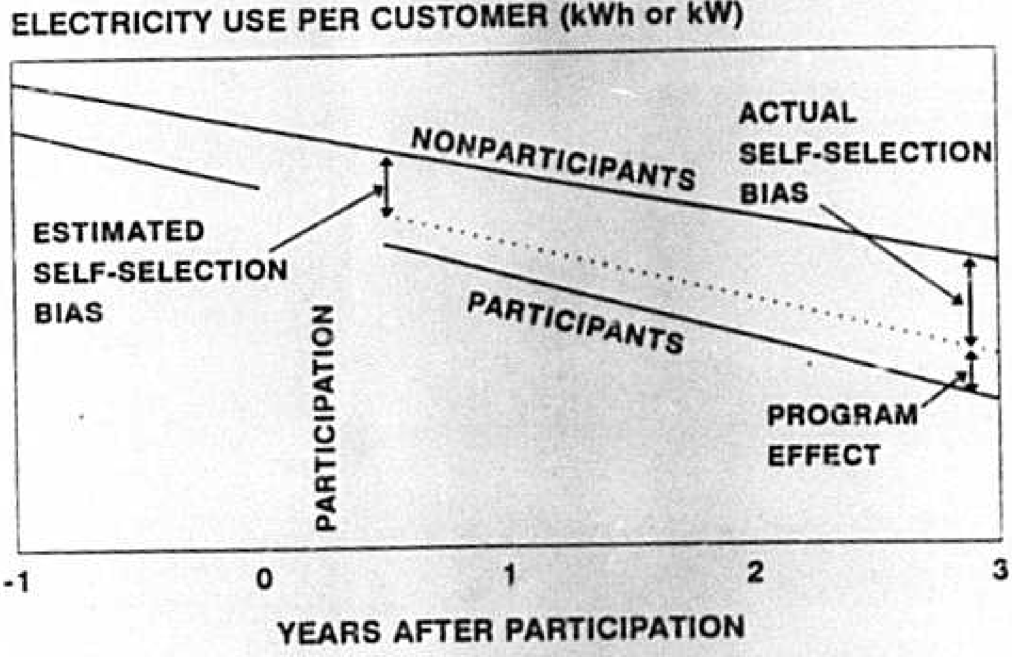


Fig. 8. Self-selection bias in pre/postmodels.

SELECTED SPECIAL ISSUES

Data Analysis and Estimation Accuracy

Some of the confidence levels and tolerance limits set as targets for DSM research are based on those developed for end-use research where the objective was to measure a consumption level, not a change over time. These precision requirem nts may not be appropriate for DSM evaluation. One criteria for determining the appropriate level of accuracy should be whether the information is adequate for making good business decisions regarding investments in DSM programs. The confidence needed for this may be different than that required for other utility investments because of the ability to monitor DSM programs periodically and the high ratio of variable-to-fixed costs of DSM programs compared to a power plant. If savings are estimated to be 10%, then is it necessary to have 90% confidence intervals with 10% tolerances, i.e., must the utility know that the true savings falls between 9% and 11%: Or might an estimate of 10% savings, where the analyst is 90% confident that actual savings will fall between 6% and 14% (i.e., a ±40% precision), might meet the criteria for good business decisions, particularly if this number can be updated frequently as additional program experience is gained. In this case, a requirement that savings estimates achieve a ±10% precision seems high. Also, if program costs are low, the program may be cost-effective at 6% savings. Ultimately, the specification of accuracy required for decision making will be decided by the utility and other interested parties.

Multiple-Analysis Methods

Multiple analysis methods are frequently used for evaluation. For example, experimental designs sometimes include engineering estimates, metering data (whole-premise or end use), and billing data analyses (CMP 1990; Northeast Utilities 1991a and 1991b; the New Jersey Conservation Analysis Team 1990). The use of multiple methods can encompass different approaches. Three are discussed below.

- **Triangulation** involves the use of more than one approach to estimate the impacts of a DSM program. See the subsection on Combinations of Data Sources in Chapter 3 and Hirst (1991) for more information.

- **Leveraging data** differs from triangulation in that it uses data from one method as input to another method to increase the precision of the impact estimates. This approach uses data to reduce the variance in the parameter being estimated. Difference and ratio estimates are examples of leveraged approaches in which data from two analyses are combined. For example, in SAE methods, the engineering estimates of program or measure savings are used as inputs to the statistical model (EPRI 1991a): see Chapter 3 for additional discussion of this topic.

- **Prior information** on the variable of interest (e.g., program savings) is used as a starting point. The uncertainty around this estimate is quantified from previous studies or expert judgement. This approach uses this information as an input to a Bayesian-estimation framework. Information on the impacts of DSM programs and measures can come from prior evaluations, engineering models, and expert judgement. The use of this information with a systematic approach to updating and adapting existing estimates based on new evaluation data could usefully augment conventional estimation approaches. In some cases, this may be a successful primary estimation method when the objective of the evaluation is to update existing estimates. This approach is discussed further later in this chapter.

The use of multiple methods and data sources allows the evaluator to use the available information in a cost-effective manner. Information in a DSM evaluation can be leveraged through a number of methods.

- Using end-use metering information in engineering models for calibration, validation, and simulation of a hypothetical baseline (Kaplan, Jones, and Jansen 1990; EPRI 1991b)

- Using engineering relationships to specify statistical models (King and Gavelis 1990)

- Using engineering estimates and/or metering to augment statistical estimates (EPRI 1991a and 1991b; Empire State Electric Energy Research Corp. 1990)

- Combining population and sample information

- Combining expected or forecasted savings with statistical models by using statistical audit procedures or using SAE estimates
- Using prior information in the form of savings estimates and distributions (EPRI 1991a)

Procedures That Accumulate Information Over Time

One concern with current methods is that each evaluation essentially starts over, ignoring the results of previous evaluations. This procedure may be appropriate for medical research where falsely rejecting the null hypothesis carries an extreme penalty. However, the appropriateness of this assumption for DSM evaluation is an open question. The use of prior information in a Bayesian framework assumes that information should be cumulative. On this basis, the researcher can start with existing estimates and then revise these estimates based on new data.

The key requirement for a Bayesian analysis is the establishment of an acceptable and unbiased "prior estimate" of impacts or consumption levels. While this requirement may be a difficult one to meet, these techniques might usefully augment traditional estimation approaches. The use of prior information as constraints, as initial estimates of savings, or in the form of a likely distribution of savings may all be useful. These approaches assess the likelihood of observing a drawn sample, given the conditionally assumed prior estimates. The use of these methods may be one way to achieve the levels of accuracy that some analysts are hoping for in evaluation. Software is now available that allows the analyst to start with an existing estimate that is bounded by wide CI (EPRI 1991a; Pollard 1986; Schmitt 1969).

Free Drivers and Comparison-Group Contamination

The issue of free drivers is increasingly recognized as important in DSM evaluation (Chapter 8). Several approaches for addressing this issue have been suggested. One approach is to change the estimate of baseline energy use from which the impacts are measured. For example, it may be appropriate to use a historical baseline from the early years of the program. Free drivers are more likely to be a significant problem for programs that have been in existence for several years and have achieved high participation levels. If baseline energy use for nonparticipants is taken from one of the earlier years, then the concern about nonparticipant consumption being influenced by the program is reduced. Because it can be assumed that some energy-efficiency improvements were likely to occur without the program, this approach typically would provide an upper bound on savings. However, if the time span is only two or three years, this improvement could be expected to be minor.

A second approach involves the use of survey methods to determine whether nonparticipants have changed their energy use as a result of the program, as discussed in Chapter 8. Nonparticipants could be queried as to whether their behavior has changed because of the program. Trade allies could be queried to determine whether they believe the market for the DSM actions promoted by the program has moved in such a way that nonparticipants are impacted (CMP 1990).

While free drivers are one form of comparison-group contamination, several other concerns regarding contamination have been raised as well. These concerns include whether or not to include in the comparison group customers that participated in DSM programs other than the program being offered. A second issue is whether to exclude from the comparison group customers that participated in the same program prior to the analysis period (Chapter 6). In some instances, these customers cannot be cost-effectively identified and excluded from the comparison group. The question is whether this inclusion poses a potentially important bias. In most cases, this will not be an important problem. If a change model is used, then a number of factors will reduce the potential influence of these factors on impact estimates. Information should be collected on all programs in which the customers have participated to allow for testing of the significance of this factor on estimates, but the use of a pooled cross-sectional time-series experimental design should reduce any bias this might cause.

End-Use and Whole-Premises Metering Versus Billing Data

The finer the temporal and end-use resolution in the consumption data, the better the analysis (Chapter 7). The statistical problems and methods discussed earlier in this chapter are equally applicable to end-use, whole-premise, and billing data. Using difference- and ratio-estimation methods to increase precision is more relevant with end-use data because of their ability to reduce the required sample sizes to achieve a given level of precision.

There have been only a few metering studies that have been specifically designed to estimate the savings resulting from DSM programs (Northeast Utilities 1991a; Violette 1990). Pre- and postparticipation metering are required to accurately capture the change in consumption. Several utilities have incorporated end-use metering projects in evaluation plans. As these studies are completed, considerable information on the applicability of metering will be available.

Engineering Methods

Engineering approaches can provide information on equipment and system performance characteristics and operation profiles of measures installed through the programs. These techniques have seen widespread application in DSM-program impact evaluation as well as in program planning and screening (EPRI 1991a).

Early comparisons between impact-evaluation methods using consumption data (e.g., billing histories) gave estimates that were substantially lower than the engineering estimates. However, recent work using engineering methods based on onsite inspections after installation of DSM measures show much closer agreement between methods.

Engineering methods can serve several useful functions in impact evaluations.

- **Engineering estimates may provide independent, stand-alone estimates of program impacts.** Engineering methods provide a quick method of developing before- and after-the-fact estimates, as well as ongoing monitoring of program impacts. As an evaluation method, however, engineering assumptions and estimates need to be benchmarked against billing-data analyses and end-use metering studies.

- **Engineering estimates can verify statistical methods.** Statistical approaches can produce implausible estimates for certain programs, often because of data-quality problems. Engineering estimates can help assess plausibility and provide backup estimates.

- **Engineering methods are sometimes the most cost-effective method.** Engineering methods are inexpensive and can serve as a primary method when the value of information does not justify more-expensive statistical and end-use-metering approaches.

- **Engineering estimates can provide time differentiation of impacts.** Statistical estimates are limited by the temporal disaggregation of available consumption data, often a one-month billing period. Valuation of impacts in benefit/cost analyses often requires estimates of consumption impacts differentiated into on-peak and off-peak periods and load-shape impacts by hour of the day. Savings from billing analyses can be allocated to these time periods with engineering models.

- **Engineering approaches can estimate savings for other fuels.** Billing data for natural-gas or fuel-oil consumption may not be available to an electric utility, yet the program may influence the consumption of these fuels. In some cases, it may be more cost-effective and expedient to rely on an engineering estimate.

- **Engineering methods can offer measure-specific resolution.** Statistical approaches may be constrained in determining impacts by individual measures. The available data for a billing analysis may be on a package of measures, and multicollinearity may prevent the identification of measure-specific impacts. Engineering estimates can allocate aggregate savings to individual measures or combinations of measures, including interactive effects. For example, a DSM program may offer a package of hot-water conservation measures. Engineering methods can apportion the estimated total savings for the package to individual measures (e.g., low-flow showerheads, pipe insulation, and water-heater wrap).

CONCLUSIONS

Performing impact evaluations will continue to challenge evaluators, who will apply methods that draw from many areas of expertise. These areas include survey research, sample design, statistics, market research, consumer theory, economics, and engineering. Future DSM evaluations will likely be characterized by several factors:

- Increased emphasis on discrete-choice modeling of DSM decisions. Such analyses develop better estimates of program impacts through improved baseline usage estimates. Also, these methods provide valuable insights that can be important for program marketing and understanding customer decision processes.

- More end-use and whole-premises metering. The costs of metering are declining, and experimental designs are being developed to use this information more efficiently.

- Nearly universal analysis of electricity billing data, even if additional metering is being conducted. Billing data are readily available, and the additional information required for even the most detailed analyses of billing data are generally collected through customer surveys. Showing how monthly energy consumption changed is the starting point of most evaluations. Additional methods can be used to explain why this change is or is not an accurate estimate of program impacts, but most decision makers want to see this basic calculation.

- Use of multiple approaches, in which the output of one method is used as the input to another method. Examples of these approaches include difference estimators, ratio estimators, and SAE estimates.

REPORTING AND USING EVALUATION RESULTS

Steven Nadel

INTRODUCTION

This chapter discusses what evaluation material to report, how to report that information, and how to use the evaluation results developed in the other chapters. This chapter emphasizes the orderly and complete reporting of all major types of evaluation information and the diverse uses for evaluation data, both internal and external to the utility conducting the program.

REPORTING EVALUATION RESULTS

Sometimes evaluation results are reported in a single document; other times several discrete reports are issued. The most common reporting arrangements are separate process- and impact-evaluation reports on each program or annual reports that summarize all available information on a utility's DSM programs.

In addition to basic process- and impact-evaluation data (Chapter 1), other types of evaluation data are important and useful, including data on participation rates, program costs, measure life, and free riders. Often, these data are included as part of evaluation reports, but these data are frequently omitted from such reports entirely. Much of these data are available in databases that track program activity. Because these data are continually compiled, they can be reported at almost any point, although annual reporting is most common. Table 16 provides a checklist of the information that should be included in comprehensive evaluation reports; most of these topics are discussed below.

Table 16. Key ingredients of a comprehensive evaluation report

Table of contents	Program costs
Program description	Program energy and load reductions
Evaluation method	Measure life
Participation	Cost-effectiveness
Number of participants	Process-evaluation results
Participation rates	Summary and recommendations
Free riders	

Program Description

Each report should include a description of the program being evaluated to provide a context for reviewing the results. The description may be a brief summary (with citations to a more in-depth description), or a complete description may be included in the evaluation report or in an appendix to the report.

The description should include: a history of the program (e.g., when it began and ended); program goals, including the target audience; an indication of whether the program is a pilot or full-scale program; the program's components (marketing, incentives, information, technical assistance, etc.); the program's operations (audit, rebate application, etc.); and changes made to the program since previous evaluations or program descriptions were published.

Evaluation Method

Every evaluation report should discuss the methods used. For process evaluations, the description of methods may be brief, indicating who was interviewed, how these people were selected, what questions were asked [commonly a copy of the interview guide(s) is included in an appendix], what other data were used and their sources, and who conducted the interviews and other evaluation activities.

For impact evaluations, the description should cover sample sizes, sample selection, comparison-group selection, data-cleaning rules (what criteria were used to eliminate or modify poor-quality data), data attrition (how many customers in the original sample were eliminated from the analysis and for what reasons), and the analysis procedures used. This information allows the reader to assess the validity of the approach and the results obtained. When discussing data attrition, summarizing the results in a table is often useful (see Table 17). In discussing the method, explaining why particular approaches were selected is also useful. The method write-up often includes citations to other sources that describe particular evaluation approaches in detail.

The methods used to track and calculate other program data should be reported, including how the participation rate, free-rider proportion, direct and indirect costs, and energy and demand savings were defined and calculated. Because different evaluators use different approaches and definitions for key terms, it is important to specify which approach and definitions were used (Berry 1989 and 1990; Nadel 1991; Hirst and Sabo 1991).

Participation

Data on the number of participants is used to assess the overall size of a program and to calculate indices, such as participation rate and cost per customer. The number of participating customers is typically calculated in one of three ways:

Table 17. Data attrition in the low-income segment of CMP's Weather Shield and Attic Attack Program[a]

	Number of cases	
	Participant group	Comparison group
Original database	1,904	442
No billing data available	274	---
Master meter; less than 300 days of data; vacancies; duplicate accounts	226	33
Greater than 50% change between pre- and postparticipation years	86	36
Remaining observations	1,318	373

[a]This table shows the original sample size for the participant and comparison groups and the number of cases that were lost at each stage of data cleaning, including missing data ("No billing data available"), unusable data ("Master meter, ..."), and data that distort the analysis ("Greater than 50% change ...").

Source: Stucky et al. 1990.

- The number of applications processed (including multiple applications submitted by a single account)
- The number of unique account numbers participating in the program (including multiple account numbers used by the same home or business)
- The number of unique customers participating in the program

While the last definition is probably the most useful, it is frequently impractical to implement because most record-keeping systems cannot identify all accounts associated with a particular home or business. The second definition is easier to calculate because identical account numbers are generally easy to find. For this reason, it is likely to be the preferred definition for most applications. Use of the first definition should generally be avoided because many programs encourage customers to participate more than once, which makes impossible the calculation of how many eligible customers have not participated in a program.

Three other important issues in calculating the number of participating customers are whether to count pending applications or only completed applications, whether to count all participants or only participants who are not free riders (see Chapter 8), and whether to count audit recipients or rebate recipients in combined audit/rebate programs (not all audit recipients apply for a rebate). Because some pending jobs are never completed, counting only completed applications is usually best. Generally, in calculating participation, no adjustment is made for free riders; instead, the free-rider proportion is tracked separately. For combined audit/rebate programs, the number of audit recipients and rebate recipients are usually counted separately because both figures are useful.

Evaluators should also categorize participants by market segment, such as customer electricity use (peak demand or annual kWh use), business or house type, income or economic activity, geographic region, and the DSM measures installed. Categorizing participants makes it possible to compare participants to eligible nonparticipants to see which groups participate and which do not.

Participation rate, the number of participating customers divided by the number of eligible customers, indicates the proportion of customers served by a program and, by subtraction, the proportion that have been missed and hence need to be the target for future efforts. The number of eligible customers may be all residential or C&I customers, or it may be a subset (e.g., C&I customers with peak demand greater than 500 kW).

Sometimes, the number of customers eligible for a program is large, but a much smaller number is specifically targeted. In such cases, evaluators should report two participation rates, one based on eligible customers and one based on targeted customers. In calculating the number of eligible customers, the same rules should be used as are used for calculating the number of participants (Berry 1990; Nadel 1991; Hirst and Sabo 1991).

For some programs, participation rates may be based on factors besides participating and eligible customers. For equipment rebate programs, such as motor, ballast, and refrigerator programs, the participation rate may be defined in terms of pieces of equipment that received rebates relative to annual sales of that type of equipment in a utility's service territory.

Both annual and cumulative (since program inception) participation rates should be reported. In addition to reporting participation rates, interpreting the rates obtained in terms of the utility's short- and long-term goals for the program is useful. For example, is the program on schedule?

As noted in Chapter 8, estimating free riders is an imprecise science. Therefore, in reporting free riders, reporting a range and which point within the range is most likely (for use in calculations that require a single-point estimate) is often appropriate. Because free-rider rates are likely to vary from measure to measure, evaluations of programs that promote multiple measures should present free-rider estimates by measure.

Program Costs

Information on program costs is needed for budgeting as well as for determining program cost-effectiveness. Costs associated with a program fall into three general categories: direct utility costs, indirect utility costs, and nonutility costs (Berry 1989; Hirst and Sabo 1991). Direct utility costs include monies paid to customers and contractors for the purchase and installation of DSM measures. Indirect utility costs are all other utility costs, including funds for inhouse staff, marketing, evaluation, consultants, etc. Nonutility costs are the monies paid by customers and other involved parties, *after* crediting any utility incentive.

Direct costs are usually the easiest to obtain because they are commonly tracked in the program's database. Indirect costs may be tracked at the individual program level or may be tracked at the department level. When costs are tracked at the department level, they should be allocated to individual programs. In reporting indirect costs, indicating which costs are included (e.g., staff, marketing, evaluation, etc.) and which, if any, are not included is important. Nonutility costs (e.g. net customer costs to purchase and install a measure) have traditionally not been tracked by utilities. However, because these costs are needed to calculate cost-effectiveness from the societal perspective, some utilities now collect these data. Customer costs are commonly obtained in one of three ways: requiring customers to submit invoices, asking customers to report their costs, and compiling estimates of typical customer costs. The first method is accurate, but places an extra burden on customers. The second method is less accurate but less burdensome. The third method usually involves detailed data collection for a sample of customers and extrapolation to the entire population of participants. This method can be accurate if samples are carefully drawn and accurate information is obtained from customers. In reporting data on customer costs, reporting the method used to compile the data is important.

Program Savings

Data on kWh and kW savings are needed for analysis of program savings and cost-effectiveness. However, kWh and kW figures are difficult to interpret unless savings data are referenced to preprogram-consumption data. For example, savings of 1000 kWh/year, commendable for a customer using 3000 kWh/year, are trivial for a customer using 1,000,000 kWh/year. The most common way to reference savings is to calculate savings as a percentage of the average preprogram, whole-building electricity use or demand of participating customers.

In reporting savings, it is important to distinguish and report total savings and net savings (Chapters 1 and 4). Energy savings may be based on engineering estimates or on statistical analysis of metered energy use and demand (Chapter 4). In reporting savings estimates, noting which method(s) was or were used to develop the estimates is important. Also, for statistical analyses, reporting information that indicates how accurate the savings estimates are likely to be is useful. Perhaps the most useful measure of statistical accuracy is the CI. Unlike other statistical measures, such as significance levels, *t*-values, or SD, CIs are easy for nonstatisticians to understand. When reporting the CI around a savings estimate, the confidence level (e.g., 90%) should also be reported (Table 18).

Demand savings vary by the hour of the day and the day of the year (Chapter 7). Rather than reporting 8760 hourly values for each year, evaluation reports usually provide more limited data, such as demand savings coincident with a utility's summer and winter peaks.

Table 18. Electricity savings from the Northeast Utilities EnergyCHECK program[a]

Annual electricity use and savings (kWh/ft^2)	Participants (n = 54) Mean use	95% CI	SD	Nonparticipants (n = 75) Mean use	95% CI	SD
Preimplementation use	12.73	±3.06	11.49	12.81	±4.35	19.23
Postimplementation use	12.48	±3.12	11.70	13.79	±4.53	20.02
Change in use	-0.25	±0.88	3.31	0.98	±0.76	3.38
Net program impact (difference between participants and nonparticipants)	-1.23					

[a]This table shows net and total savings and includes data on sample size, CI, and SD. Although the table does not show percentage savings, data on preprogram electricity use is provided, permitting percentage savings to be easily calculated. Data is reported on a per square-foot basis because participant buildings were slightly larger than buildings in the comparison group.

Source: Peters, Oswald, and Horowitz 1990.

Some utilities report maximum demand savings or average demand savings. Maximum demand savings is the largest demand reduction achieved by a measure, although this reduction is not necessarily achieved at the time of the system peak. Average demand savings are kWh savings achieved by a measure divided by 8760 hours (the number of hours in a year). For example, replacing a 40-W lamp with a 34-W lamp has a maximum demand savings of 6 W. If 70% of the lamps are on at the time of the summer peak, coincident summer peak savings are 4.2 W [6(0.7)]. If the lamp operates for 3500 hours each year, average demand savings are 2.4 W [6(3500/8760)].

Many cost-effectiveness models separately examine program costs and savings by season (e.g., summer, winter, and spring/fall) and by time of day (e.g., peak, off-peak, and shoulder periods). These models generally require estimates of *average* demand savings during each of the periods involved. In reporting these values, which months, days, and hours are included in each period should be indicated.

Even more information on demand savings can be provided in graphic form. For example, demand savings can be plotted as a function of time of day or day of the year. Time-of-day graphs typically provide hourly information. Often, graphs for several days are provided (e.g., for the peak day and for an average day during the summer and winter). Graphs of this sort, while useful for all programs, are particularly useful for load-management programs because such graphs show not only the peak demand savings but also the increases in demand during other periods (Fig. 9).

Reports of kWh and kW savings should indicate whether savings are at the customer or generation level, and if the latter, what transmission and distribution losses and reserve

In reporting demand savings, summer and winter coincident peak savings should generally be reported. In addition, maximum savings, average annual savings, and average savings during specific seasons are often useful.

Measure Life

The life of DSM measures, an important variable for determining program cost-effectiveness, is also critical for capacity-planning purposes because, when measures wear out, they must be replaced by new resources (Chapter 6). Estimates of the effective life of each DSM measure promoted by a program should be reported along with the source of each estimate.

Program Cost-Effectiveness

Cost-effectiveness is usually assessed by comparing the benefits and costs of a program. If the ratio of benefits to costs is greater than one, a program is cost-effective. To standardize these calculations, the California PUC and the California Energy Commission (1987) developed several tests to calculate the benefit/cost ratio for DSM programs. These tests examine program cost-effectiveness from the perspectives of participants, nonparticipants, the utility, all parties (called the total resource cost test), and society (the same as the total resource cost test with the addition of quantified externality costs). Krause and Eto (1988) explain and demonstrate the use of these tests.

Evaluation reports should include the results of these five tests for each program analyzed. Benefit/cost ratios can be calculated on a cumulative basis (since program inception) or for the most recent year. In addition to providing benefit/cost ratios, evaluation reports should include an explanation of the key assumptions used in the calculations, such as discount rates, avoided costs, and environmental externality values.

Costs per kW and kWh provide a quick approximation of program cost-effectiveness, although these indices cannot substitute for a detailed benefit/cost analysis. Costs per kW are calculated by dividing program costs by kW savings. To be most useful, these calculations should include both direct and indirect costs. Savings should be estimated for the time of the system peak. Costs per kWh, also called the cost of conserved energy, are generally calculated on an average basis over the life of the measures installed. This is done by levelizing, over the average measure life, the annual program costs with a discount rate equal to the utility cost of capital. The cost of conserved energy is the annualized payment divided by the kWh savings in one year (Meier 1982). For example, if a program costs $1 million, saves 4 million kWh annually, has an average measure life of 10 years, and a 6% real cost of capital, then the annualized program cost is $135,868, and the cost per kWh is 3.4¢/kWh ($135,868 divided by 4 million kWh).

Costs per kW and kWh can be compared to reference values based on utility avoided costs (¢/kWh) and to the cost of new power plants ($/kW). Whenever utility DSM-program costs are substantially less than these reference values, the program is likely to be cost-effective. Of course, a definitive estimate of cost-effectiveness requires a full benefit/cost analysis. In particular, using the ¢/kWh index does not give any credit for kW savings (and vice versa).

Programs that are not cost-effective based on kW or kWh savings alone may be cost-effective when both benefits are included. The costs for peak-clipping programs should be compared to the cost of new peaking power plants (in $/kW), while the costs of energy-efficiency programs should be compared to the cost of new baseload power plants (in ¢/kWh or $/kW).

Process-Evaluation Results

Process-evaluation results may be reported in a separate process-evaluation report, or they may be one element of a larger evaluation report. In either case, process-evaluation results should be reported in the context of other evaluation findings. That is, where appropriate, the process evaluation should address issues raised by other evaluation activities and vice versa. In this manner, the results of different evaluation components, when taken together, provide a complete, integrated evaluation (Chapter 9).

Summary and Recommendations

The summary may be the most important section of an evaluation report because utility executives and PUC commissioners may read only that part (Exhibit 7). In addition, many people will read the summary before deciding whether to read additional sections of the report. Thus, the summary must include key evaluation findings and recommendations. The summary should be written for senior management and should strike a balance between brevity and thoroughness. Recommendations and action items should be prominently featured. The summary should interpret evaluation findings in light of short- and long-term program objectives. For example, how do participation rates compare to estimates made at the beginning of the year? In addition to providing a written summary, summarizing data on the program (e.g., costs, savings, and participation rates) in a single form may be useful to readers so they can readily find the data they need (Hirst and Sabo 1991).

Recommendations from an evaluation usually fall into two general categories: ways to improve the program being evaluated and follow-up work to provide more and better evaluation data. In reporting recommendations, each recommendation should be highlighted so that managers can quickly find the recommendations that interest them. The accompanying text should describe and summarize the rationale for each recommendation, and also identify who (e.g. program manager or evaluation supervisor) should be responsible for reviewing and implementing each recommendation. If evaluation recommendations are acted upon before the evaluation report is completed, the report should indicate the disposition of each recommendation. More often, the disposition of recommendations will have to be reported in subsequent evaluation reports, and the disposition of previous recommendations reported in the current evaluation report.

USING EVALUATION RESULTS

Evaluation results can be used for both internal (within the utility) and external purposes. In this section, we discuss many of these uses and conclude with a brief discussion on internal versus external reporting.

Exhibit 7. Evaluation summary for the C&I lighting program of NEES (1988)

This summary describes the program, participation, savings, costs, and cost-effectiveness. These findings are interpreted in light of long-term goals. Several recommendations are also discussed. This summary has only a few weaknesses: (1) participants are reported in terms of number of applications, not number of customers; (2) participation rate is not reported; and (3) the type of energy savings reported is not described (in this case, net generation-level savings based on engineering estimates; demand savings are coincident with the summer peak).

> The C&I Lighting Rebate Program promotes the installation of energy-efficient lighting products through dealer incentives. Dealers are paid rebates for sales of qualifying products to customers of NEES's affiliated retail companies. Products currently eligible for the program are energy-efficient fluorescent lamps, ballasts, fixtures and reflectors, and high-intensity discharge and compact fluorescent lamp retrofits. 1987 was the start-up year for the program.
>
> In 1987, 104 dealers (out of a target group of 179 dealers) submitted 1449 rebate requests totaling $516,696. Of the money rebated, 36% went to fluorescent lamps, 11% went to fluorescent ballasts, and 52% went to fluorescent fixtures. The majority of rebate requests came from 18 dealers. Nearly half the customers listed on rebate requests come from the large C&I revenue class, despite the fact these customers account for only 6% of NEES's C&I customers. For the program to be successful, more dealers need to participate in the program, and participating dealers need to handle an increased volume of rebates. One major target of these efforts should be small and medium-sized C&I customers.
>
> Energy savings from equipment sold under the program in 1987 are estimated to be 1.3 MW and 4.8 GWh per year. These savings figures have been adjusted to eliminate estimated savings by "free riders" — customers who would have purchased eligible equipment even if the program were not offered. Free riders are estimated to account for approximately 45% of fluorescent lamp rebates and 50% of fluorescent fixture rebates.
>
> The program is functioning smoothly. Dealers appear to be having little difficulty with the rebate forms, and rebate checks are being issued promptly. Overall, a survey of dealers indicates that the majority of dealers are very satisfied with the program. Only 6% of the dealers surveyed were dissatisfied with the program.
>
> Based on the program's performance in 1987 and based on projections of future program participation levels and costs, the program is projected to just meet its target of 57 MW of peak demand savings. Under these projections, the program's cost-benefit ratio is estimated to be 0.21. Even if participation levels are half of projections and the number of free riders increases significantly, the program still has a cost-benefit ratio of 0.50 or less. However, if participation levels are low or free riders increase, the program will not meet its peak demand savings target. In order to allow for these contingencies and still meet the program's demand savings goal, it may be desirable to add additional types of products to the program (beyond those added in December, 1987) and/or to take steps to increase savings from products currently eligible for the program.

Internal Uses

Internally, evaluations are used for many purposes, as illustrated in Table 19.

Table 19. Internal uses of evaluation results

Use	Example(s)
Program modification	Process evaluation shows how to attract new market segments to a program. Impact evaluation shows that some measures save less energy than expected.
Resource planning	Impact evaluation shows the likely future effects of programs to be included with power plants in utility resource plan. Process evaluation identifies role of trade allies, leading to creation of a new program.
Load forecasting	Evaluation data is used to strengthen behavioral relationships within forecasting models. Impact evaluation results are used to adjust forecasts for future effects of utility DSM programs.
DSM bidding and performance contracting	Impact evaluations of utility-run programs provide a yardstick against which to compare bids. Evaluations suggest which programs a utility should run itself and which ones to operate through contractors.
Rate design	Time-of-use response to interruptible and other rate programs can be used to design new rate programs.
Market research	Process evaluations identify differences among market segments in participation, satisfaction with the utility, its programs, and other characteristics.

Program Modifications. Evaluation results can be used to modify programs. For example, as a result of an evaluation of a pilot program, a utility might decide to implement a full-scale program, incorporating program modifications suggested by the evaluation. Process evaluations often include recommendations for program modifications that can be quickly assessed and implemented.

For example, BPA ran a pilot program in which industrial customers submitted proposals for partial funding to install DSM measures. Measures were required to have a payback period of three years or more. A process evaluation found that the three-year payback criterion exceeded the investment threshold of many firms. The evaluation also found that, because of the competitive nature of BPA's selection process, many firms were unwilling to invest in proposal preparation unless projects were already under consideration at the plant, which meant that a high proportion of free riders was likely. Also, the two-month period during which proposals were accepted did not correspond to the capital-budget cycle at most plants (Peters and Gustafson 1987). BPA subsequently changed the incentive to a ¢/kWh-saved basis and allowed proposals to be submitted at any time (BPA 1988).

In addition to adopting short-term recommendations, evaluation results (e.g., participation rates and savings per customer) should be compared to program expectations. If results are significantly less than expectations, further analysis is needed to decide whether the initial expectations were reasonable and if program performance can be improved. For example, an evaluation of BPA's Institutional Buildings Program found that savings were 40% less than prior engineering estimates. Likely reasons for the discrepancy were identified, and suggestions were made as to how the program could be modified to improve future savings (Keating and Blachman 1987). Several of these recommendations were incorporated into the program, and many of the remaining recommendations were incorporated into new programs subsequently developed by BPA (Keating 1991).

Evaluation results should also be compared to studies of the conservation potential within a utility service-area. For example, if a conservation-potential study indicates that 500 MW of savings are available from motor-efficiency improvements but motor savings after two years of program operation are only 20 MW, then additional steps may be needed to tap more of the potential. This happened in NEES's initial motor program; low savings relative to the potential led to a complete revamping of the program (Stout and Gilmore 1989). Following the program changes, participation rates increased substantially (Nadel 1990).

Utilities should compare evaluation results with results from similar programs operated by other utilities. Comparing results for several different indices is useful; such indices include participation rates, percentage kWh and kW savings, free-rider proportions, ratio of indirect to direct costs, cost per kWh saved, cost per kW saved, and the ratio of evaluation estimates of energy savings to engineering estimates. If these indices are calculated with the procedures discussed above and in Nadel (1991) and Hirst and Sabo (1991), they are usually comparable from program to program. If an index value for a particular program is better than that for other similar programs (e.g. high participation rate), it may indicate that a program is going well. If an index for a program is below average, it indicates that additional research is needed, first to see if there is a reasonable explanation for the poor index value and second to see if program improvements are justified to improve the index value.

Resource Planning. Evaluations of existing programs often provide useful information for planning new programs and for preparing long-range resource plans. To illustrate, Nadel and Tress (1990) developed long-range DSM plans for three New York utilities. For each of the 21 programs examined, estimates of participation rates, utility and customer costs, savings per customer, and measure life had to be made to compute energy savings, costs, and cost-effectiveness. Wherever possible, these estimates were based on the results from evaluations of successful utility programs.

Forecasting. Evaluations can be used to improve utility load forecasts by providing data needed for forecasting models and by helping to estimate the impact of DSM programs. Most utilities incorporate DSM measures into their forecasts using a two-step process. First, baseline energy-use and demand forecasts are prepared that incorporate the effects of DSM actions taken by customers in the absence of utility programs. Second, the impacts of utility-operated programs are specifically modelled and subtracted from the baseline forecast.

In computing baseline energy use, future efficiency trends in the absence of utility programs must be estimated, including the impacts of market forces and government efficiency standards. Data collected for evaluations can help in this regard, particularly data on free riders and energy-use trends among nonparticipants. For example, baseline data collected to evaluate an appliance rebate program can be used to estimate the energy consumption of new residential appliances. Similarly, baseline data collected for a commercial new construction program can be used to estimate the energy use intensity (kWh/ft^2) of new buildings. These values are required inputs for many forecasting models. Some forecasting models include consumer-choice algorithms in which adoption of DSM measures is modeled as a function of measure costs, savings, and current stage on the technology-diffusion curve. Evaluation results can be useful in estimating these model coefficients.

DSM Bidding and Performance Contracting. Many utilities contract with energy service companies (ESCO) and other contractors to provide energy savings. In such cases, the utility must determine whether the energy savings and cost estimates provided by prospective contractors are reasonable. Also, the amount of energy savings actually achieved must be estimated, particularly if payments to contractors are based on kW and kWh savings. DSM evaluations can help with both tasks. Evaluations of existing programs provide useful yardsticks for assessing the reasonableness of bids and proposed contracts. and past evaluation can suggest cost-effective ways to measure the energy and demand reductions provided by contractors.

Rate Design. Data collected on electricity demand during different hours of the day and different seasons can help in the design of time-of-use and interruptible rates. An evaluation of NEES's Stand-By Generation Program (which provided payments to customers to operate emergency generators during peak periods) provided insights on how to improve program participation and savings. These results were subsequently used to revise the rate schedules for the company's Stand-By Generation and Interruptible Service Programs (Nadel, Cress and Ticknor 1989). Similarly, Southern California Edison used performance data on storage cooling systems to design special time-of-use rates to encourage such systems (Nadel 1990).

Market Research. Evaluations collect information on the market segments targeted for the program, including data on customer characteristics and preferences. For example, Southern California Edison maintains a database of market-research information on its largest customers. As evaluations are conducted on programs serving these customers, additional information for the database is likely to be produced.

External Uses

Regulators use evaluations in their review of DSM plans and performance. Intervenors in the regulatory process use evaluations for the same reasons. In addition, DSM-program planners and implementors from other utilities use such results to learn from the experiences of others.

Regulators may want to review evaluation plans prepared by utilities for reasonableness (are the approaches proposed in line with modern evaluation practice?) and thoroughness (are all needed data included?). Such reviews help ensure that the information regulators need will be collected, reducing the chances that disagreements on procedural issues will hamper the recovery of DSM-program costs. For these reasons, the Connecticut and Massachusetts PUCs require utilities to submit detailed evaluation plans for approval.

Regulators might require utilities to collect and report specific types of evaluation data. For example, the Massachusetts Department of Public Utilities developed a specific format for annual reports of DSM results (Hirst and Sabo 1991). Regulators might even specify specific evaluation methods. For example, the New York PSC directed its utilities to work with commission staff to develop a standard method to evaluate three DSM programs mandated by the PSC (XENERGY 1990).

Of course, PUCs will review the results of evaluations. Typically such reviews are ad hoc; commissioners and commission staff read evaluation reports and ask questions. One approach to a more systematic review is for regulators to compare evaluation results to utility short-term and long-term goals, estimates of the available conservation resource, and results of similar programs operated by other utilities. As discussed earlier, using several indices can make this task easier (Nadel 1991).

Occasionally, when regulators are dissatisfied with utility evaluations, or when they want an independent review of evaluation results, they may hire an evaluator to conduct such a study. For example, the Wisconsin PSC has done this on a few occasions (Nichols et al. 1990).

Finally, regulators can use evaluation results to help guide cost-recovery determinations for DSM programs, including recovery of program costs, lost revenues (reductions in revenue collection due to reduced kWh sales caused by DSM programs), and financial rewards for successful implementation of DSM programs (Chapter 1). Evaluations are generally the source of program costs and savings estimates upon which cost recovery is based.

Typical intervenors in PUC proceedings include public interest groups, state consumer counsels, and large industrial customers. Each of these parties will selectively choose evaluation results that reinforce its own case. For example, in a recent proceeding before the District of Columbia PSC, witnesses for the local consumer council, citing a recent review of DSM-evaluation results, argued that Potomac Electric Power Company's proposed DSM programs had significantly lower participation rates and savings than the most successful programs offered elsewhere in the country and that the utility should substantially increase its targets for DSM savings. Witnesses for the utility, citing the same study, argued that proposed programs had participation rates and savings near the industry average, and hence its targets were reasonable. In this case, use of evaluation findings allowed the different parties to agree on the facts (the utility's targets were near the average in the industry, and not among the most ambitious) but to disagree on what actions to take. The agreement on the facts made it easier for the PSC to make a decision in the case. The commission encouraged the utility to exceed its proposed savings targets, but did not require it to do so (District of Columbia PSC 1991).

Evaluation results can be used by other utilities, ESCOs, and state energy offices. Unfortunately, obtaining evaluation data from others is not always easy, because most evaluation results are not widely distributed or publicized. Several compilations of evaluation results have been published (Nadel 1990; Blevins and Miller 1989a and 1989b; Northeast Region Demand-Side Management Data Exchange 1989; Keating and Hicks 1990; Energy Program Evaluation Conference 1991). These publications can be useful but suffer from limited availability (the Northeast Region Demand-Side Management Data Exchange data are generally available only to members), limited coverage (only some programs are included), missing data, and time lags between updates.

Internal vs. External Reporting

Because evaluations are used both internally and externally, the question inevitably arises whether internal results should be reported externally. A number of considerations affect this determination. First, in many jurisdictions, once results are put in writing, even in an internal memorandum, regulators and intervenors can request the information during the next regulatory proceeding for which the data are relevant.

Second, if information is withheld, such as disappointing results for a particular program, and regulators discover it, the ramifications are likely to be far worse than if the results were voluntarily reported in the first place. An open sharing of information with regulators is likely to improve a utility's credibility, resulting in long-term benefits to the utility.

Based on these considerations, internal results should generally be reported externally. The advantages of external reporting generally outweigh the disadvantages. Although most results should be reported externally, they should first be reported internally, starting with staff who worked on the program that was evaluated and then proceeding to utility management. Such a successive reporting strategy allows errors to be caught before results receive wider distribution. Also, a successive review process allows each reviewer to consider the findings in private before making a public response. With this type of process, program staff and utility management are less likely to be defensive about evaluation recommendations and are more likely to accept recommendations that make sense.

Also, some information is particularly sensitive, with little to be gained by public reporting. Information that falls into this category can include the names of employees, customers, and trade allies who made mistakes. Such information is of little concern to others, and by publicizing the information, much animosity is generated that may cause more harm than good. This type of information should generally be reported verbally only to those people within the utility who can constructively use the information (McRae 1990).

CONCLUSIONS

A thorough evaluation requires collecting, analyzing and reporting a substantial amount of data, including process, impact, and tracking information. This information should be presented in a clear, well-organized manner, so that those who use the report can easily find specific information. When information is reported, the meaning and derivation of each data element should be clearly described (for example, that the savings reported are net savings at the customer level, as determined with conditional demand analysis).

While much is known about how to report and use evaluation results, additional work still needs be done. Some of the definitions for specific types of evaluation data are imprecise (e.g., program participant and eligible customer) and could benefit from standardization. A recent study on terminology and reporting formats should help substantially in this regard (Hirst and Sabo 1991).

When evaluation data are reported, they often appear in utility or consultant reports that are not widely distributed. To make this data available to a wider audience, an index of evaluation reports should be prepared and regularly updated. Logical organizations to undertake such an effort include the national laboratories or EPRI. Furthermore, existing databases of evaluation results should be expanded to include more programs and data and should be regularly updated. To address this need, Lawrence Berkeley Laboratory is planning to develop and regularly update a database of DSM results (Association of DSM Professionals 1991). In addition, Hirst (1990b) proposed that EIA collect and compile annual data on individual DSM programs just as they presently collect annual data on individual power plants. EIA has begun to collect limited data on utility DSM efforts, but has yet to compile data on specific DSM programs.

PERSISTENCE OF ENERGY SAVINGS

Kenneth M. Keating

INTRODUCTION

Persistence of the energy and load reductions from DSM programs is a crucial issue for program planners and a difficult one for evaluators. The planned value of savings from these programs, and hence the cost-effectiveness, depends on the continued impact of the program over the projected life of the program measures. Unless the savings from the program continue, the alternate resources deferred by the program will be needed sooner than expected. Without persistence, the DSM resource loses its long-term value. Put simply, if it is not there when you need it, it is not worth much.

In the initial development of DSM programs, persistence, a nagging issue for electric-system planners, was assumed by program operators to be reasonably constant. In many instances, experience with DSM measures was lacking, making it difficult to determine the effective life of a measure. Most planners assumed that knowing the physical life of the measures installed was sufficient to determine persistence. As DSM programs have matured and as the money spent on such programs has increased dramatically, utilities and regulators have begun to look closely at persistence. Evaluators are being asked to validate the persistence assumptions of program planners as part of integrated resource planning.

Although early evaluation reports suggested a slight decline in the net program savings of some programs (Hirst and Keating 1987), limited data supported changes in the dominant assumptions of basic persistence. Now, program and evaluation experience indicate that persistence is a pressing and difficult issue. DSM-program managers can no longer install a measure, declare victory, and walk away.

UNDERSTANDING THE ISSUES

To answer the question "Do the impacts of DSM last?", evaluators need to be clear about which impacts are being examined. Persistence has two dimensions. One dimension concerns the lifetime of the DSM measures installed by a program, the manner in which these measures are operated, or both. The second dimension concerns the overall definition of persistence. Defined at the utility level, persistence is equivalent to the long-term temporal changes in net program impacts; defined at the societal level, it is equivalent to the long-term temporal changes in total impacts. Note that the persistence of both total and net impacts includes technical and operating characteristics.

The net-program perspective requires that persistence be measured with an eye to what the participants would have done in later years had there been no program. If nonparticipants in the comparison group adopt energy-efficient practices and measures, this perspective in effect reduces the value of the savings in later years. It reflects the probable actions of the participants had the program not existed. This concept is theoretically appealing, but the results obtained with a comparison group can be confounded by spill-over effects from the program. In essence, nonparticipants may adopt program-recommended actions because of the program even though they do not formally participate in the program: see the discussion of free drivers in Chapter 8.

In considering total impacts, what others are doing is irrelevant as long as the measures in the participants' facilities continue to perform over time as they did when first installed. The key issues here are the technical performance of the DSM measures and the way that these measures are operated and maintained. These technical- and operating-performance estimates are relative to the performance of the conventional measures replaced. Consider a situation in which electricity use for a high-efficiency refrigerator increases by 5% after two years because the householders did not clean the compressor coils. If the same increase in electricity use would have occurred with a conventional refrigerator, no loss of savings is associated with the high-efficiency unit.

Figure 10 illustrates the different perspectives for a residential retrofit program. In this example, the program's net impacts decline over time as the general population adopts more-efficient measures and practices. From the "efficiency" perspective, some measures deteriorate (e.g., deterioration of weatherstripping after five years, reducing savings until it is replaced in year nine). In either case, the amount of savings is the area under the respective curves, not the first-year savings continuing for 20 years.

Technology and behavior affect persistence. These two factors interact and are often difficult to separate. Examples of technology failure include equipment malfunction, breakdown, or loss of efficiency over time. Behavioral issues include removing measures, failing to provide maintenance, or overriding controls. Technological and behavioral aspects interact when measures, such as efficient lights, are removed because they were inappropriately recommended and provide insufficient light or when the high-efficiency technology does not produce the requisite comfort. For example, people may remove low-flow showerheads if they do not like the "feel" of the shower they get with such units. It is also possible that a manufacturer could increase production *as a result of* having a more efficient operation. This behavioral response is called takeback and will erode net program impacts. Figure 11 illustrates the various threats to persistence discussed here.

Program managers should be concerned about the actual life of many commercial-sector technologies (and some residential ones also, such as low-flow showerheads and compact fluorescent lights). DSM-program planners often base estimates of measure lives on

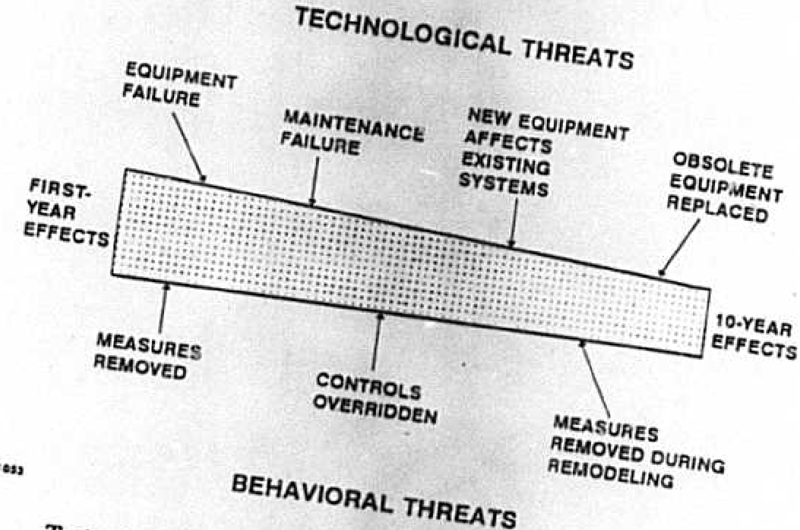


Fig. 11. **Technological and behavioral factors that affect the long-term energy savings of DSM programs.**

replaced with less efficient fixtures in a few years. Data show that commercial buildings typically undergo some type of remodelling or renovation every few years (Petersen 1990; Petersen and Sandler 1991; Puget Energy Services 1991; Kunkle and Johnson 1991; Hickman and Steele 1991).

Persistence is not an either-or question; it is a matter of degree -- how much of the impact remains over time. DSM programs install a mix of measures in a variety of buildings. Some measures in some buildings will fail or be removed. However, program evaluations deal with aggregates of buildings, and although technological or behavioral failures may dramatically affect savings for individual buildings, their effect may be limited for an entire DSM program.

Exhibit 8. Frequent changes in commercial buildings shorten effective lives for many DSM measures.

In an extensive research effort for BPA, 100 professional remodelers were asked to estimate the frequency of remodelling in different types of buildings, and which end uses were likely to be affected. In addition, the researchers visited 300 commercial buildings that had undergone remodeling or renovation to find out what had been changed, and what happened to the old equipment (Skumatz et al. 1991). The results suggested that the effective life of many common DSM measures was only half that of the manufacturer estimate. For example, lighting fixtures might be replaced every four or five years in small office buildings and restaurants.

METHODOLOGIES AND ISSUES IN PRIOR RESEARCH

Limited research is available on the persistence of DSM savings, but horror stories abound: power outages causing electric timeclocks that control water heaters to lose so many hours that the heating elements are on at system peak; floor insulation being ripped out to fight termites; or pipes freezing, cracking, and soaking the insulation in walls. DSM opponents point to these occurrences as proof of the ineffectiveness of DSM measures. DSM proponents argue that each of these glitches lost only a tiny amount of a diverse resource, compared to a loss of 800 MW caused by a bad valve at a power plant. Both sides are right. Perfect persistence is an impossibility. However, the erosion of savings may be modest. A key evaluation issue concerns how savings vary with time. A critical planning issue concerns how to design programs so that their impacts are more persistent.

Two general approaches to research on persistence have been used:

- The total-impact perspective is associated with the study of measures in place: follow-up site visits or surveys to verify continued installation and proper functioning of measures have been the emphasis in several studies.

- The net-impact perspective is associated with analysis of electricity-billing records of participants and those of a comparison group of similar nonparticipants. This approach may not work well for large C&I customers because a suitable comparison group may not exist.

Studies of Measures in Place

Persistence of efficiency impacts may be tracked by verifying the continued presence and proper operation of program-installed measures when the measures operate on a known and unchanging schedule (e.g., motors on 24-hour operation, exit lighting, and street lighting), when there are few interactions with other energy-using systems or occupants, and where the savings are well established (e.g., those of water-heater tank wraps or low-flow showerheads).

Surveying samples of participants is an inexpensive approach for tracking some measures. Central Hudson Gas and Electric conducted a short-term study of persistence of WH efficiency improvements. In its household surveys, between 87% and 94% of the measures installed were reported to still be in place up to 29 months after installation (TechPlan 1990). The measures included tank wraps, low-flow showerheads, and pipe insulation.

A two-step survey process was used to study the persistence of residential lighting measures in New England. In a process evaluation, participants reported the removal of 12% of the high-efficiency compact fluorescent bulbs within seven months of their installation (Applied Management Sciences, Inc. 1990b). A year later, NEES conducted follow-up site visits to 96 of these low-income homes. The company found that 25% of the bulbs had been removed, representing 20% of the displaced wattage. Most occupants removed the bulbs because they were dissatisfied with the quality, color, or quantity of light (Tolkin 1990). This erosion of savings was caused by the interaction of behavior and technology. These results do not necessarily mean that compact fluorescent lamps will always show a loss of savings; the results may provide evidence of potential weaknesses in program design, measure selection, or program delivery.

A more expensive two-part study was undertaken on buildings that participated in the BPA Institutional Buildings Program in Washington State. A program engineer visited 49 sites (74 facilities) three years after the buildings were retrofitted (Kunkle 1990). The second stage of the research was a telephone follow-up to the facilities one year later (Kunkle and Johnson 1991). A major advantage of the follow-up was that it tracked what happened to major systems when problems developed. In 27 facilities, there had been 41 problematic measures, centering on control systems, HVAC, and heat-recovery systems. Of the 12 problem control systems, five were disabled, three repaired, three replaced, and one was left unrepaired. Of the 12 problem HVAC retrofits, three were disabled, four repaired, one replaced, and four continued to operate inefficiently. Although the poorly performing systems were less than 25% of the systems installed (and only 5% of the lighting retrofits had problems), significant erosion of savings occurred during the first six years after installation.

Studies of measures in place face limitations. Problems arise in quantifying and extrapolating the effects on persistence when some measures are missing or when they operate less than optimally. To illustrate, if the efficient motors examined during a site visit are not being maintained properly, exactly what kWh impact should be inferred for these motors? What should be inferred for other motors installed through the program but not examined during the site visit? In addition, unless this type of verification is extended to a sample of nonparticipants, the program savings may be underestimated because no account would be taken of efficiency losses for standard equipment. Measuring persistence without a comparison group implicitly assumes that standard equipment experiences no deterioration in performance over time.

Analysis of Electricity-Consumption Data

Analysis of billing data is often associated with the net-effects perspective because it has traditionally involved the use of comparison groups of eligible nonparticipants. This approach is essential for attributing the changes in consumption to the program and for separating these effects from those caused by other (nonprogram) factors. This approach also adjusts for the effects of free riders.

This approach has been used in several studies of short-term persistence, but rarely to examine persistence for more than a few years. BPA studied three retrofit-program cohorts (1981, 1985, and 1986) for three postparticipation years (Hirst et al. 1985; Keating and Hirst 1986; Haeri 1988; White and Brown 1990). The results were mixed. The savings for some cohorts remained essentially constant, while the savings for others declined with time.

Analysis of data for three postparticipation years for participants in the Hood River Conservation Project showed a complete loss of savings. But most of the erosion was caused by the participants' switching SH fuel from wood to electricity (Schoch-McDaniel 1990). This finding demonstrates the importance of using a comparison group, which was not possible in Hood River, where 90% of the homes participated. If a comparable group of nonparticipants had been available for analysis, they might have shown the same pattern of fuel switching because of the dramatically improved economic conditions in the Hood River area.

A parallel study of the persistence of the demand impacts (kW) of the Hood River Conservation Project indicates that the effect on weekday and annual peaks did persist extremely well for three years (White et al. 1991). These Hood River results suggest that the performance of the retrofit measures persisted largely unchanged; net savings declined because of operating changes (primarily the shift from wood to electricity for heating).

In the commercial sector, a comparison group is essential to control for the frequent fluctuations in building energy use. In particular, changes in business cycles can have large effects on energy use (Keating and Oliver 1991).

Two long-term studies of persistence used billing data from participants and a comparison group. Both involved residential retrofit programs in the Pacific Northwest, one by Seattle City Light (Sumi and Coates 1988) and one by BPA (Horowitz et al. 1991). Each followed samples of participants and nonparticipants for six years, and both weather-adjusted electricity consumption with PRISM (Fig. 12). Two results are evident: the trend in energy savings is downward, and the decline is erratic (i.e., net savings vary from year to year). The

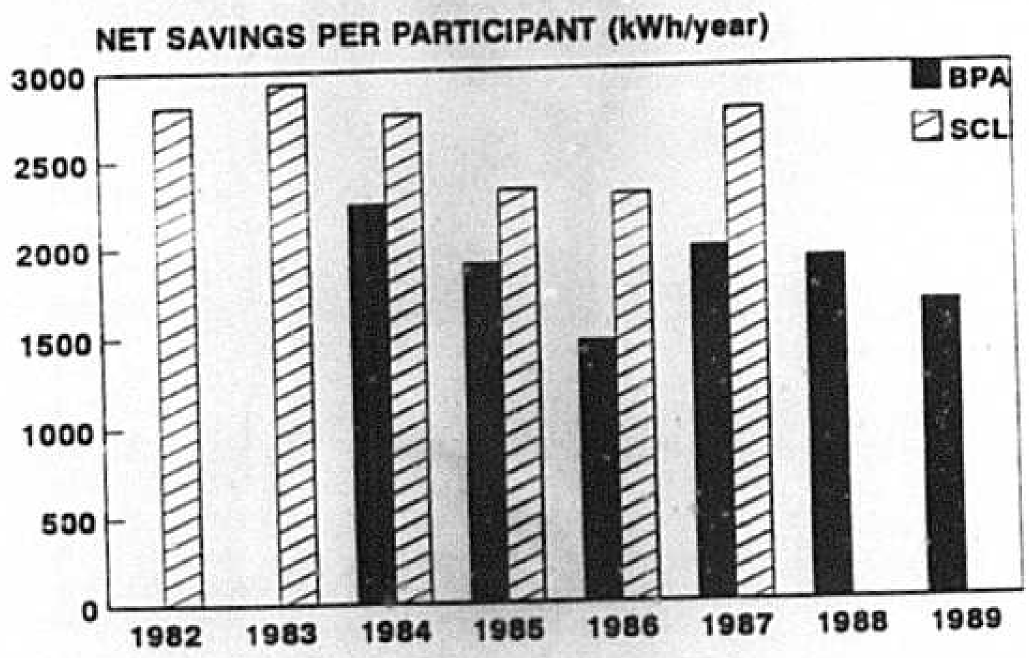


Fig. 12. Net electricity savings for several years produced by two residential retrofit programs in the Pacific Northwest.

average decline in savings for the last five years compared to the first year amounts to 7% for the Seattle program and 21% for the BPA program. These studies indicate substantial persistence, but also some erosion of savings. The eighties were characterized in the Northwest by unprecedented electricity-price increases for the first half of the decade, followed by declining real prices. Much of the erosion of savings was caused by reduced consumption by the comparison groups, about 2000 kWh during the six years studied, primarily because of electricity-price increases.

The reports from BPA and Seattle show that these data can be aggregated and analyzed many ways. For example, evaluators in Seattle were asked to provide a forecast of future changes in electricity savings, not for any particular cohort, but in general. Therefore, their report includes a weighted average savings for all cohorts for each year (i.e., the average savings in 1983 for homes weatherized in 1981 and 1982; the average savings in 1984 for homes weatherized in 1981, 1982, and 1983; and so forth). This practice reduces the year-to-year fluctuations within cohorts and shows an average erosion of savings of 27% during the six-year period (Sumi and Coates 1988).

These studies exemplify two problems with the use of billing analysis to study persistence: attrition bias and the confounding effects of free riders and free drivers. Sample attrition occurs when households move, billing data are lost, or nonparticipants become participants. Attrition bias occurs if the remaining cases are systematically different than the cases that were lost. The BPA participant and nonparticipant samples suffered an attrition rate of 55%. Six years after retrofit, they had only half as many cases with consumption data as they did one year after retrofit for that cohort. Because the first-year savings for the group with six years of postprogram data was much smaller than that for the original sample, attrition bias was apparent. The study used survey data to correct for the small portion of the bias that could be quantified.

The Seattle study, on the other hand, shows how to avoid attrition bias in a longitudinal study. Seattle reported only 10% sample attrition because Sumi and Coates (1988) used a retrospective approach. That is, they started their research in 1987 and selected their samples of participants and nonparticipants from among those for whom they had continuous billing records back to the year before program participation for the participants and to 1980 for the nonparticipants. This meant that all the people who moved or joined the program late were already excluded. Because this study examined the same households for the entire period, the results are not affected by attrition bias. On the other hand, the samples of customers that Sumi and Coates were left with were special groups and were surely not representative of all customers. Their approach, however, should be considered where possible.

Controlling for free riders and free drivers in persistence research from the net-program perspective is daunting (Chapter 8). The effect of free riders on net program savings is usually accounted for by the use of a comparison group to estimate what the behavior of the participants would have been without the program. However, what if the comparison group takes action *because of* the program? This behavior, a spillover effect of the program, is called a free-driver effect.

participants, chose to participate in the program. Nevertheless, sample attrition and attrition bias will remain, particularly in the commercial sector.

An, as-yet untried alternative is an econometric approach. A dynamic econometric model that defines the relationships among electricity use and electricity prices, measures of economic activity, and program participation could be developed. This model could then be used to predict what consumption would be in future years if energy efficiency were held constant and the actual values of electricity prices and other explanatory variables were entered into the model. If the consumption observed in the out-year was higher than predicted, then the efficiency effects could be said to be eroding, and the effect could be quantified.

Finally, many projects that are in idiosyncratic settings (e.g., industrial-process retrofits) will have to be studied on a case-by-case basis to measure persistence. In these situations, the evaluation may involve more than measurement. It may form a partnership with maintenance and monitoring functions so that persistence problems found in the evaluation provide feedback to the maintenance staff.

CURRENT AND PLANNED RESEARCH

Several persistence studies are under way or in the planning stage. The California Energy Commission sponsored a workshop in early 1991 to review ways to study the issue state-wide. The BPA study (Skumatz et al. 1991) reviewed above is the first step in a major study of persistence in the commercial sector. It is limited to studying only one threat to persistence, the removal of measures caused by remodellings, renovations, and business changes. If energy-using systems (such as lighting, HVAC, and motors) are frequently removed, the expected lives for efficiency measures will be much less than their physical lives. Further research will be done to determine the extent of maintenance failures, measure failures, and the importance of consumer training. In general, the amount of research devoted to persistence is expected to grow. EPRI (Hanser 1991) and the New York State Program Evaluation Task Force (Cummings 1991) each have made persistence a high priority.

PROGRAM IMPLICATIONS AND CONCLUSIONS

Concerns about persistence should affect programs in both planning and implementing cost-effective programs. Evidence continues to accumulate that program planners have generally overestimated savings (Nadel and Keating 1991). Empirical research raises questions about the validity of using manufacturers' claims for physical measure lives as a basis for projecting persistence. Planners need to become more conservative in estimating the effective life of program-induced savings.

If managers want long-term effects, they need to be concerned about the longevity of the results. One approach is to focus program resources on new construction, remodelling, and renovations. This approach allows utilities to pay a lower incentive based on the incremental cost of the efficiency improvement (rather than the total cost, which might be paid for a retrofit). It also installs DSM measures when the length of time to the next remodelling is the longest.

A second suggestion is to make sure that the installed equipment is commissioned properly – that controls work, that systems work together rather than out of synchronization, that air flows are balanced, that the measures provide the desired amenity level, and that the occupants know how to operate the measures properly. This will reduce the likelihood that occupants will override control systems, remove measures, drop out of the program, or "bad mouth" the program to other potential participants.

A third approach is to provide continuing education and maintenance. Many times when occupancy or staff change, the new people do not know how to deal with the efficiency measures and may ignore their potential to save energy (Komor et al. 1989). Follow-up can be accomplished by making maintenance contracts a part of the measure package, by flagging vulnerable technologies for yearly follow-ups, and by providing general education for building owners and maintenance workers. Performance contracts for third-party measure installations that base payments on savings over time can be effective for some classes of buildings and measures, but it can be expensive, and it removes the customer contact from the utility. Nevertheless, it is potentially effective for targeted segments of the DSM market or as an interim approach until the utility can establish its own expertise.

Persistence is a genuine problem of undetermined scope. Its effects on cost-effectiveness, program planning, and resource reliability are clear. It is now time to address persistence in earnest.

USE OF LOAD RESEARCH IN EVALUATION

George Fitzpatrick

INTRODUCTION

Load research is the collection of electricity-usage data through a metering device associated with an end use, a circuit, or a building. The goal of this activity is to produce cost-effective, statistically sound, detailed databases for resource planning, evaluation, and rate-design studies.

Load research has been employed by electric utilities since the 1960s to better understand the characteristics of electric loads, the timing of their use, and the amount of electricity consumed by users. During the 1970s, load-research activities expanded as the needs for the data increased beyond those for traditional cost-of-service studies. As early as 1976, load-research information had become a direct input into at least one utility's load forecast (New York Power Pool and the Empire State Electric Energy Research Corporation 1976). Even earlier, class load research had been employed in the construction of weather-normalized class-by-class system peaks (Long Island Lighting Company 1972). The enactment of the Public Utility Regulatory Policies Act of 1978 spurred utilities to make additional investments and kept manufacturers of load-research equipment and consultants busy meeting the demand for load data.

There are two generic types of load-research information. Direct load-research information is the measured class or end-use demand (kW), energy (kWh), or reactive power (kvar). The sampling, collection, editing, and analysis of these measurements are the primary activities of direct load research.

Indirect or corroborative load research is the data that the load researcher uses to explain and understand direct load-research measurements. These data include appliance-saturation data, demographics, DSM-process-evaluation data, weather data, equipment-nameplate data, audit data, etc.

Load research is expensive. Therefore, care must be used in starting or expanding such efforts. The design and conduct of a load-research program requires knowledge of research objectives, sample designs, and common load-research needs across DSM programs.

LOAD-RESEARCH PROGRAM ACTIVITIES

Load-research programs have eleven major activities (Table 20), each of which is discu below. While these activities are discussed in a linear fashion, the activities are, in f continuous and often take place in parallel, particularly as priorities change and r requirements arise. A successful program emphasizes planning and design while striving maintain the efficiency of load-research operations.

Table 20. Key activities associated with load research

Program planning	Meter placement and replacement
Sample design	Installation record keeping
Sample selection and validation	Meter operation and maintenance
Recorder selection	Data collection
Meter quality control and quality assurance	Data validation and editing
	Data analysis

Program Planning

In this activity, overall program objectives are defined; the data requirements (both direct and indirect) are specified; the common needs of utility users of load data are identified; program timelines are determined; and staffing, equipment, and software requirements are specified. This activity is a vital first step, but it is also an activity that should be revisited frequently during the conduct of a load-research program.

Planning should include investigating the experiences of other utilities with load-research programs, including visits to view their load-research operations. If visits can be arranged, they should include discussion with the load-research manager and DSM-program managers using load-research data, the viewing of field installations, discussions with installers and maintenance personnel, and observation and discussion of procedures for recruiting customers, installing equipment, data collection, and data editing.

There are other important sources of information on load research. The Association of Edison Illuminating Companies conducts seminars on load-research methods for utility staff with minimal experience as well as regional conferences for advanced practitioners. The American Public Power Association Rates and Load-Research Committee may also be contacted for valuable guidance. In-house seminars conducted by consultants can address specific topics, such as analysis software, editing strategies, and development of applicable statistical skills.

Given the significant lead times required to purchase and install load-research equipment (a minimum of four to six months), the evaluator must identify load-research data requirements well in advance of when load research begins. Load-research planners find it useful to have evaluators develop a timeline showing when different types of data are required. With this information, the load-research planner can anticipate purchasing and installation requirements. For example, if a DSM program has as its focus summer-peak reduction, then evaluation plans must be finalized by December of the previous year. Knowledge of the evaluation cycle of programs that require load research is essential for competent planning.

Sample Design

Sample design was discussed in Chapter 3. But an important distinction must be drawn between a load-research sample and a sample to be used for evaluation. In the case of the load-research sample, the researcher is interested in predicting the time-varying load or class of loads. For example, the goal might be to predict the diversified demand of air conditioners with a 95% accuracy. If the average air conditioner load is 3.25 kW and the average duty cycle at the 3:00 PM peak is 0.6, then the average predicted load would be 1.95 kW, and the sample size would be chosen to make this accurate to ±0.09 kW.

In contrast to the load researcher, the evaluator usually is interested in the *change* in load. Thus, if the program manager's goal is to reduce AC load by an average of 10%, then the sample chosen has to be large enough to predict an average reduction of 0.195 kW with an accuracy of ±0.009 kW, assuming the estimate is to be accurate within 5%.

A second difference between a load-research sample and an evaluation sample concerns external validity (see Chapter 2). The load researcher may only be interested in the behavior of a class of customers, such as users of central AC. The evaluator is likely to be interested in understanding the air-conditioner usage patterns of subgroups of air conditioner customers or in generalizing the results (external validity) to a larger group of customers. Thus, a simple random sample of air-conditioner users may not be sufficient for an evaluator who needs to have the sample of air-conditioner users stratified by key variables, such as total monthly energy consumption, business type, size of business, household income, size of household, type of structure, size of structure, or air-conditioner sizing.

Billing, load-research, and survey data can be used to help identify appropriate sample sizes. For example, if the target for a load-control program is customers with central AC who use more than 1500 kWh/month in the summer, sample-size requirements can be estimated for certain cycling strategies with estimated diversified demand curves developed from load-research data. End-use data are invaluable for developing baseline information for all DSM programs.

A key challenge is to develop an overall plan that integrates samples at as many points as possible to achieve the most economical pattern of metering. Sample designs can be developed by creating new samples or enhancing existing samples. For example, a utility's in-place class load-research program might be modified and/or increased in size to allow for the metering of a representative sample of one or more key end-uses. This might require adding a transponder to take advantage of channel capacity in an existing recorder, adding more channels to an existing recorder, or adding households to the sample. This strategy optimizes the use of available equipment and takes advantage of other data collected from load-research customers.

Some important cautions about this approach should be noted. First, the integrity of the load-research sample must be maintained while providing an adequate evaluation sample. Existing load-survey customers may have an unmetered load available, but their customer profile may not fit the sample design requirements, thus leading to bias. Upgrading the metering and acquiring additional information from the customer may require further intrusions upon customers' time and privacy, causing customers to drop out. Thus, gains in economy may be offset by the need to deal with sample bias.

Sample Selection and Validation

Sample selection is the translation of a complex sample design into a practical set of activities leading to the installation of metering devices. This activity has several outputs:

- The primary samples (lists of customers to be contacted) for all programs
- Multiple backup samples for all programs
- Primary and alternate customer lists for meter installations
- In some cases, customer lists for personal or mail surveys

The responsibility for control of the sample should be clearly specified and should be known to all program participants. Any question or problem regarding a sample should be referred to that person.

A customer chosen to be in a sample may have to be dropped for any of several reasons: a difficult or impossible installation, no telephone line accessible from the metering point, a customer's refusal to participate, or a customer not having the appliance of interest.

The randomness of a sample must be ensured when a customer must be replaced. Installers tend to want to complete an "easy" or "next-door" replacement installation rather than to complete an installation that preserves the statistical integrity of the sample. The sample-design coordinator should keep a backup selection file and should specify any replacements on a customer-by-customer basis until the full sample is completed.

The problem of sample replacement is even more acute in the case of a stratified sample. Purists may argue that random replacement within a stratum is sufficient, while pragmatists may argue that the replacement sample point should have characteristics that are as close as possible to those of the sample customer being removed. The following method may satisfy both perspectives:

- When drawing the initial sample, draw ten separate samples, each with the same number of respondents as specified in the original design.

- With the original-sample design parameter(s), validate each sample by comparing means and SDs.

- Select the best "fit" sample, by stratum, as the installation sample.

- Take the remaining samples that fall within acceptable confidence limits (e.g., 99% to 90%) and combine these samples into a master backup list maintaining the stratification and the design criteria (e.g., where the design criterion is peak-month billed kW, segment the "master backup" by both stratum and 10-kW increments).

- When replacing sample points, randomly draw from the master backup sample from within the appropriate design segment.

This method ensures that the overall central tendency, variance, and distributional shape are closely maintained over the course of the load-research experiment. A variation on this concept that provides even greater fidelity to the original design entails the use of appliance survey data to develop a replacement matrix. This method provides greater matching precision while maintaining statistical integrity.

Recorder Selection

Equipment selection is driven by a number of criteria. These criteria may include the type of measurements to be made (e.g., whole-building versus end-use data), the frequency with which measurements must be made, and the future uses of the equipment.

Whole-building meters can be placed in an existing standard watt-hour socket. These devices provide both billing and load-research data. These meters can be read by a handheld recorder, a handheld device aimed at the meter while driving by, or a modem. These meters can contain logic and sufficient memory to store many days of readings taken at intervals ranging from a minute to weeks.

Multichannel end-use meters are usually installed independent of the billing meter. They are capable of taking data from 2, 4, 8, or 16 channels. Transponders at the metering points may be hard-wired to the recorder or may communicate with the recorder with some form of power-line carrier and the building wiring. The recorder usually contains sufficient memory to store up to several days of data taken at intervals of one minute to several hours. The data stored in the recorder is usually captured by the use of a handheld recorder, a modem that calls a master station periodically over the customer's telephone lines, or a modem connected to a leased phone line.

Which of these options to use depends on the trade-offs between data requirements; capital costs: installation costs; operational costs (such as the use of a meter reader, the customer's phone line, or a leased line); analysis costs; and restoration costs to the customer's premises.

For example, using a power-line carrier may have lower installation and restoration costs than hard-wiring transponders, but it will likely have higher capital costs. Whole-building metering is generally less costly than multichannel metering but may require more sample points to obtain the same accuracy.

An important issue is the coordination of master-station software with recorder selection. Most utilities now have inventories of recorders from different vendors. Different vendors use different communication protocols. If the vendor for the master-station software is different from that of the recorder, then the master-station software supplier should communicate with the recorder supplier about appropriate modifications to the software. The products of some independent suppliers of master-station software can communicate with a variety of recorders, thereby integrating recorders from different vendors into a coherent system.

If software is to be developed (or modified) in-house, then early communication with the recorder supplier will facilitate development. Early attention to this detail may suggest some easy modifications to recorder output data or data formats to meet particular in-house analysis requirements.

Care in recorder selection can produce significant savings for the load-research program. A common problem is the mismatch between current meter inventory and the requirements for the next load-research experiment. Plans for the purchase of new equipment should allow future experiment cycles to use equipment and installation and removal labor efficiently.

Spare recorders should be available at all times. With the popularity of stratified samples, the loss of a few recorders, especially in the same stratum, can render entire periods of data statistically unsound. Utilities considering joint load-research efforts should investigate the possibility of integrating samples and sharing spare equipment.

Meter Quality Control and Quality Assurance

Meter quality control and quality assurance is an essential ongoing activity that ensures the reliability and the cost-effectiveness of the load-research activity. Reliable metering reduces the need for oversampling and lowers costs. The level of reliability depends on whether or not the data will also be used for billing.

Meter problems have several potential sources, including metering equipment that is faulty upon arrival from the factory, equipment that is improperly installed, faulty record keeping (for example, misrecording information about the location of a meter or channel information), and failure of meters in the field for hardware or environmental reasons.

The utility should establish a test bench and test all metering equipment supplied by manufacturers. All of the equipment to be used at a site should be tested together. The site-specific settings should be made and tested while the equipment is on the test bench. Appropriate labels (barcodes) should be attached to the individual devices. For example, if an air conditioner, a water heater, and the total household load are to be monitored, then the channels should be preassigned, current transformers and transponders associated with the channels should be preset, and the devices should be labeled. A standard protocol should be used for assigning channels to end-uses. This practice will help reduce errors, minimize problems in the field, and aid trouble shooting. Everything should be double-checked and recorded, and then the equipment should be boxed for transport to the site.

Equipment performance should be monitored constantly. Each new installation should be polled as soon as is practical after installation, and the resulting data should be checked to see if the unit is functioning properly. Most master-station data-collection software interrogates metering devices and reports power outages, malfunctioning transponders, or high rates of communication errors. These reports should be reviewed every time the meters are polled. Range checks should be performed on the data. The data should also be reviewed to identify channels where values are always "0" or where values are valid but constant. Every instance of invalid data or suspected equipment malfunction should be followed up to identify and catalog the causes. Equipment performance should be discussed frequently with the field staff and their supervisors. These procedures can lead to the detection and correction of manufacturing and installation problems before significant data loss occurs.

The researcher has to seek the assistance of the most qualified field installers and meter-maintenance personnel. The metering and billing departments should be involved in decisions about metering devices and the planning of metering operations. Without such consultation, the priorities of field personnel are likely to be set by other organizations, and the quality of load-research data will suffer. Regular meetings should be held with field personnel to review equipment performance and to reinforce their understanding of the importance of the load-research program to the utility.

Meter Placement and Replacement

Meter placement and replacement are continuous activities, but periods of intense effort occur at the beginning and end of each research cycle. Proper staffing and effective planning can minimize the cost of this activity. Predefined placement and replacement procedures ensure consistency and efficient installation while maintaining the overall statistical quality of the experiment.

The resource requirements for meter placement depend on the type of metering equipment, the familiarity of the field staff with the equipment, the capability of staff to do maintenance and repair, and the type of communications. The meters for a sample cannot be placed overnight. Mobilization of the necessary staff must be well planned. If hard-wiring is required and/or installation contractors are to be used, then the customer must be informed of this. Contractual arrangements must be in place to ensure the availability of the premises for a timely installation (even though descriptive information may have been given to the customers and an authorization for the installation was obtained). Contractual instruments to be used with customers should be kept simple. Good customer relations dictate the inclusion of an offer to correct any complaint or to remove the installation if the customer so desires. Resolution of complaints should be handled by the utility rather than by a contractor.

Spare equipment must be on hand and in good repair to permit efficient and timely replacement of malfunctioning units. The data-collection software should be able to document any unusual equipment outages (e.g., during replacement) and the reinitialized values of any data stored in the accumulators of the replacement meters.

Installers should treat customers' premises with care. The customers' premises must be restored to their initial condition upon completion of the program. Lack of attention to this detail can irreparably damage relations with customers. Arrangements with contractors should explicitly include restoration of customer premises, and an inspection of the final result should be conducted, preferably with the customer in attendance. Removal of equipment requires as much attention as does installation.

Installation Record Keeping

Keeping installation records is a very important but often overlooked aspect of load research. Many a utility has had to send someone to each installation to retrieve a vital piece of information needed by analysts or to determine if any of the equipment at the site is of a type recalled by a vendor. Many crews have spent hours trying to locate a piece of equipment that was "in there." More than one program manager has wondered if a particular set of failures is specific to a particular model or lot of equipment.

These examples illustrate the need for a meter-installation tracking system. Such a system needs to include the model, lot number, and serial number for each piece of equipment and its utility identification number. Having similar information about component parts is also useful. The settings for each piece of equipment should be recorded as should installation information, such as date installed, date serviced, reason for service, and name of installer.

Installers should make a map of the location of each device installed at each location. Alternatively, the installers may photograph the installation and surrounding area with identifiable features. This map or photo will help to locate a device when it is to be maintained or retrieved. Maps and photographs can be scanned into a database so that the information can be maintained and viewed electronically.

Meter Operation and Maintenance Activities

The level of company involvement in maintenance will depend on the particular operation and maintenance capabilities of the meter-maintenance staff and the type of equipment to be used. Diagnosis and/or repair of logic circuits should not be left to the company computer hacker. If the load-research program is experiencing equipment problems, the program director may want to meet with the maintenance staff and the supplier to ensure that equipment is checked upon installation and that any limitations on in-house maintenance are identified. Contracting with the supplier to maintain equipment may be required, and if so, the supplier's ability to provide timely and adequate services or replacements should be checked. A storage room full of equipment waiting to be repaired is not only an inefficient use of capital but potentially a problem affecting the quality of data provided by a load-research program. Maintenance of communication equipment may be entirely separate and may not be covered by arrangements for metering-equipment maintenance. A warranty may not be a suitable substitute for equipment repair if the program must be slipped a year while awaiting warranty resolution of equipment problems.

Data Collection

Data collection may be accomplished in several ways. Onsite collection and remote polling with the telephone are the most popular methods. The selection of the most appropriate medium involves assessing the trade-offs among cost, reliability, and program requirements. It is also a function of the geography and the rural/suburban/urban mix of the population.

Onsite data collection typically is done by using a hand-held microcomputer interfaced to a recording device through an optical port. Onsite data collection has a number of advantages, including eliminating coordination problems with the telephone company, the opportunity to visually inspect the recorder site and surroundings at the time of data collection, and sophisticated technology to permit error-free data collection.

Recent breakthroughs in hand-held microcomputer technology have enhanced the attractiveness of this method. Some units use standard microcomputer operating systems and literally guide the meter reader through the data-collection process. Incorrect identification numbers, incorrect visual meter readings, and incorrect "time in" and "time out" observations are eliminated. Further, the ruggedness of these devices ensures safe storage and transport of data to the utility's computing system and master file. The main drawback is the cost of dedicated employees. However, given proper planning and the relative ease of the data collection, this operation can be performed part time by meter personnel or by dedicated full-time employees adhering to a flexible data-collection schedule. Load-research recording devices and the portable data-collection equipment now have sufficient memory capacity to ensure that data will not be lost if collection schedules are not strictly followed.

Telephone data-collection methods are probably most cost-effective in unusual situations, such as distant or dangerous locations. Telephone polling permits frequent communication with metering devices and presents opportunities to reduce data losses. Unfortunately, many companies do not take advantage of this feature because they do not screen the data obtained from load-research programs in a timely fashion.

Load-research equipment typically features redundant data-collection methods. For example, units with telephone modems may contain optical ports for manual data collection. Data-collection costs vary depending upon the method selected. Telephone meter reading costs approximately $300 per recorder per year, while onsite manual readings through optical ports will cost about 50% more. However, cost is clearly not the only consideration, and given the cost of lost load-research data, reliability and redundancy should be equally stressed.

Data Editing and Validation

Data editing is divided into two types. Housekeeping editing is the minor editing that involves adjusting start and stop times; summation of total kWh consumption over the appropriate period; and correction of improper IDs, meter multipliers, etc. Load-data editing involves identifying and replacing missing or bad data and correcting data for outages.

Editing involves a combination of human expertise and efficient software. Depending on the storage algorithm, as much as 3.5 megabytes of data can be produced per day by 200 four-channel recorders recording data at 5-minute intervals. These data cannot possibly be visually inspected. Thus, software must be used to evaluate data quality.

Sometimes, load-research data collected for a DSM evaluation is also used for customer billing. This creates special problems for the load-research editing function because two sets of data for the same customer, one for billing and one for the evaluator, must be developed. The billing application requires rapid turnaround of edited data, so the procedures for modification or replacement of missing or damaged data must be formalized.

Validation of load data is an audit function that ensures that the data entering the master file is consistent with data that has gone before. Validation is an important aspect of data collection and sets the stage for any editing that may occur.

Since the introduction of solid state electronic load-research recording devices, the incidence of data loss has dropped dramatically. During the days of magnetic tape recorders, typical month-to-month data loss was 10% to 15%. With the maturation of electronic devices, typical data loss may average 3% or less. However, even minimal data losses can be critical for stratified samples. Therefore, the utility should develop decision rules for editing data. Where sufficient sample sizes exist and the samples have been meticulously maintained to preserve the original objective of the sample design, minimal editing may be the best course of action. However, if extensive editing is required, automated data editing systems are available to aid the process.

How to handle data if an outage occurs at a customer site has always been subject to differences of opinion. There are three basic approaches. The first is to simply insert zeros in the appropriate time slots. The second method is to remove the record from the analysis for the outage period. The third is to construct a hypothetical record for the outage period. While some may argue that outages are a normal part of a utility's business, it is probably better to recognize that every sample customer represents hundreds, perhaps thousands, of fellow customers and that the outage is not indicative of what is occurring on the system. Further, if the evaluator is interested in the impact of a device, such as an efficient air conditioner, then representing the data with zeros does not help to meet this objective. The best course appears to be to simulate the data or to remove that part of the case from the analysis.

Outages not only influence data for the period of the outage but also for a period following the outage. An early-morning outage might cause an electric water heater to run for an extended period during the late morning. Thus, it may be necessary to modify or remove data beyond the period during which the outage occurred.

Data Analysis

Both primary and secondary load-research data are required by DSM evaluations. Taken together, these data form the basis for drawing conclusions from the impact and process evaluation (Fig. 13). Here we mention a few considerations not covered in Chapter 4.

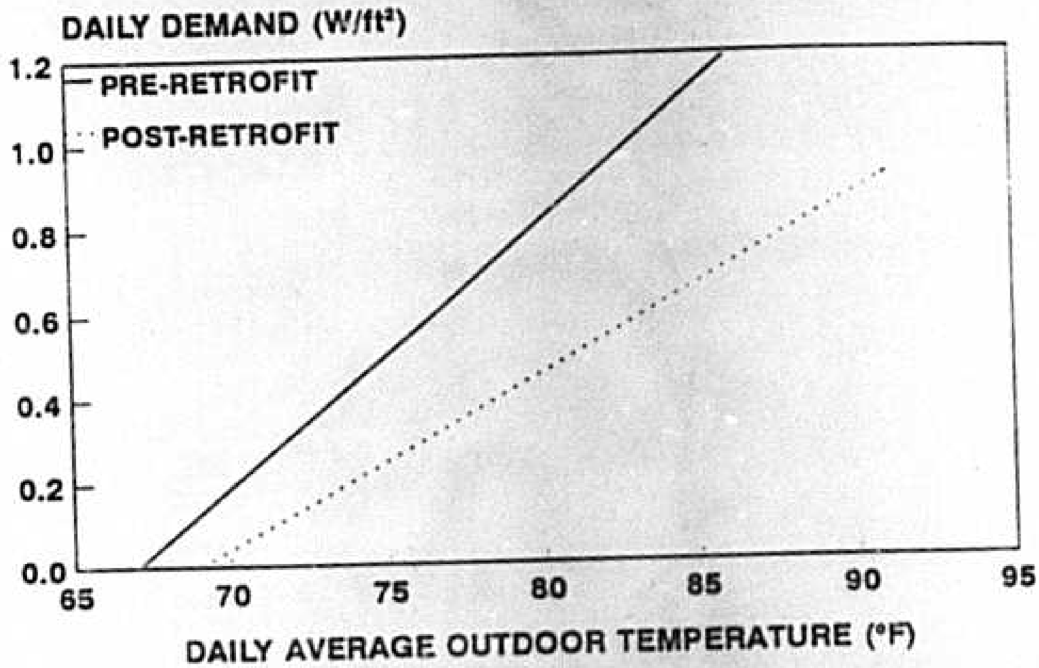


Fig. 13. Daily electricity demand for air conditioners in Austin, Texas, based on 15-minute load data collected at 14 homes. The air conditioners in these homes were replaced in late 1987 with high-efficiency units. These replacements cut peak demand by 38%.

Load-research programs often founder because of the failure to anticipate the requirements of data management. An active load-research activity can produce enormous quantities of data. It is not unheard of for data to be "lost on some tape," or for some preprocessing to be done several times because analysts do not know what preprocessing has been done previously, or for data to be lost because it is not adequately protected.

Every load-research activity should have a data-management plan. In addition to identifying how data are to be validated, such a plan should identify how the data are to be initially processed, whether they are to be stored by sample identification number or end-use or both, what quantities of data are to be stored, where they are to be stored, how files are to be named, how they are to be backed up, and any special requirements for securing access. A data dictionary should also be established. Finally, a data-management plan should identify the types of analysis that are to be conducted and the schedule for those analyses.

Data files and analyses should be tracked in a special database. The database should tell the location of the data, when it was collected, when it was processed, and how it was processed.

Analysis of load-research data typically takes a significant amount of machine time. For example, it can take a half hour or more to process a hundred thousand records on a fast minicomputer. The analyst will probably want to work with test samples to develop and test software before processing the entire set of records. Microcomputers and workstations can speed this work because of their friendly interfaces and the fact that they are dedicated machines. Some utilities now use microcomputers almost exclusively for this work, and specialized software packages are available for such machines.

Samples should be monitored to see how they change over time. Commercial firms add and release personnel, people take new jobs, household occupancy changes, and each of these factors affects energy use. Because metering samples generally are small, a few such changes may affect the character of the sample or stratum significantly, thereby influencing the generalizability of the data. For example, the estimated effects of a load-control program may be misleading if a local company adds a second shift and many persons in the sample change from day shift to night shift.

The analyst can take several actions to detect such changes. Billing data totals can be compared. If more than a year's monitoring data are available, the analyst can compare consumption data between years and relate the observations to any expected changes from the program. A follow-up survey of customers can reveal whether changes have taken place. If changes have occurred, restratifying the data may be necessary for meaningful analysis. The researcher should be prepared to evaluate all data for consistency before completing an analysis. Control groups must also be monitored for changes.

THE COST OF LOAD RESEARCH

Estimating the true cost of a load-research sample point is difficult. Economies of scale generally apply but may be lost as the size of a program changes and the use of "lumpy" resources, such as people and computers, changes. Metering costs can vary significantly depending on the number of end uses monitored at a sample location, local communications, and the difficulties of hard-wiring transponders. The first end-use point at a location is often costly, but the incremental costs for additional end-use points are less. Shared resources (e.g., programmers and field maintenance staff) may help mitigate the costs, but the effectiveness of this strategy depends on the availability of expertise and the disposition toward cooperative efforts within any particular company.

Tables 21 and 22 provide cost estimates for a range of customer sample sizes. These estimates contain assumptions about the number of end-use sample points per customer, which will vary from program to program and which can significantly affect metering costs. Also, a utility may engage a contractor as a way of trading off staffing and internal processing costs, particularly when smaller samples are involved.

Training costs have not been included. They may be significant for a small program but may be mitigated by the use of contractors, especially if the program is to begin small and grow.

Table 21. Assumptions used to estimate costs in Table 22

Meters	Points monitored		Installed	
Meter configuration	Total load	End-use points	meter cost (1991 dollars)	
Residential	1	2	2,300	
Commercial	1	2	3,700	
Industrial	1	3	5,000	
Annual charge rate				22.0%
Annual salaries			1991	
Staff classification			Annual cost	
Manager			40,000	
Data editor			20,000	
Analyst			27,000	
Programmer			30,000	
Field staff			26,000	
Maintenance staff			26,000	
Overhead loading rate				30.0%
Annual salary escalation				4.5%
Central equipment and facilities				
Type			Cost	
PCs, printers, etc., each			6,000	
Vehicles, each			14,000	
Communication equipment, per meter			30	
Minicomputer			Estimated	
Supplies			Estimated	
Purchased software			Estimated	
Annual charge rate				20.0%

Table 22. Load-research costs (1991 dollars) as a function of sample size

Sample size (meters):	150	450	750	1,500
Program size description				
Meters				
Residential	100	300	500	1,000
Commercial	30	100	150	300
Industrial	20	50	100	200
Staff (prorated for shared staff)				
Manager	0.5	1.0	1.0	1.0
Data editor	1.0	2.0	2.0	3.0
Analyst	1.0	2.0	3.0	4.0
Programmer	0.5	1.0	2.0	2.5
Field staff	1.0	2.5	3.5	5.0
Maintenance staff	0.4	1.5	2.5	4.0
Equipment				
PCs	2.0	4.0	6.0	8.0
Minicomputer	0.0	0.0	1.0	1.0
Vehicles	1.0	2.0	3.0	5.0
Summary of first costs, dollars				
Meters	441,000	1,310,000	2,205,000	4,410,000
PCs	12,000	24,000	36,000	48,000
Minicomputer	0	0	70,000	70,000
Vehicles	14,000	28,000	42,000	70,000
Communication equipment	4,500	13,500	22,500	45,000
Purchased software (capitalized)	40,000	40,000	70,000	100,000
Total first cost	511,500	1,415,500	2,445,500	4,743,000
First cost per meter	3,410	3,148	3,261	3,162
Annual costs (1991 dollars)				
Salaries including overhead	153,920	348,400	490,100	672,100
Supplies, telephone (estimated)	1,000	2,500	5,000	8,000
Office space	5,720	13,000	18,200	25,350
Present value of costs (20 years @ 11.5% discount rate)				
Meters	748,[illegible]06	2,221,969	3,740,032	7,480,063
PCs	18,504	37,007	55,511	74,014
Minicomputer	0	0	107,937	107,937
Vehicles	21,587	43,175	64,762	107,937
Communication equipment	6,939	20,817	34,694	69,388
Purchased software	61,679	61,679	107,937	154,196
Salaries including overhead	1,597,648	3,616,298	5,087,106	6,976,217
Supplies, telephone	10,021	25,053	50,106	80,170
Office space	57,321	130,276	182,386	254,037
Total present value	2,521,705	6,156,273	9,430,472	15,303,961
Present value per meter	16,811	13,681	12,574	10,203
Levelized total annual costs, dollars				
Annual cost per meter	2,161	1,774	1,631	1,323
Annual cost per sample point	696	570	521	422

ORGANIZING LOAD-RESEARCH PROGRAMS

The success of a load-research program in meeting the needs of DSM evaluation depends upon the effectiveness of communication between evaluators and load-research program managers. Many utilities already have a load-research section, often located in the rate department. Utilities that are just undertaking or expanding their load-research activities often have a concern about where the load-research function should be located.

Two strategies are possible. One is to have a load-research steering committee. This approach often is used for a mature load-research organization. The load-research steering committee meets monthly or quarterly. This committee is comprised of representatives from the traditional areas of load-research use within the utility (e.g., rates, distribution engineering, forecasting, and cost of service) as well as representatives from the residential, commercial, industrial, and agricultural DSM-evaluation efforts. Members present their needs for load research to the committee. The load-research staff analyzes these needs and designs studies to meet as many of the needs as possible. Some needs may be met without load research or by using other types of analysis, such as conditional-demand analysis. Formalization of the committee within the utility structure places the proper emphasis on the importance of such optimizing procedures.

An alternative approach is to make load research part of the DSM-evaluation function. This strategy is appropriate for utilities that are instituting or expanding small load-research programs. This organizational choice is appropriate for a number of efficiency and cost reasons. Many of the skills required of load-research personnel, including statistics, econometrics, sample-design expertise, and engineering analysis, are also required of evaluation personnel. Load-research personnel are constantly called upon to meet the data requests of evaluators. Direct organizational responsibility is advantageous in ensuring a timely response.

APPLICATIONS OF LOAD RESEARCH TO EVALUATIONS

From a load-research perspective, DSM programs fall into five basic varieties: load-control programs; direct installation and rebate programs; informational programs, including audits; time-of-use rate programs; and interruptible-rate programs. The value of load research varies according to the program being evaluated. The extent to which load research is used should be a function of the estimated level of net resource savings expected from a particular program and the extent to which an adequate analysis of impacts can be achieved through load research or some other less costly approach (Long Island Lighting Company 1991; Rochester Gas & Electric 1991).

For load-control programs, load research is the most effective way to assess the impact of control strategies and to measure program free ridership (Exhibit 9). The size of samples is dependent upon the extent to which the use of a particular appliance or end-use is discretionary and the type of load-control strategy employed. For example, the effect of load control on pool pumps can be reliably determined with a small number of monitoring

Exhibit 9. Load-control programs often require load-research data for evaluation

PG&E evaluated its residential air-conditioner-control program with meters that recorded air-conditioner electricity use every 30 minutes for a sample of about 200 participants (Malcolm 1989). These load-research data were collected for control days (when the units were turned off from 2 pm until 6 pm) and for noncontrol days (Fig. 14). Noncontrol days were chosen for their similarity to control days on the basis of outdoor temperatures. These data show a substantial reduction in demand at the time of PG&E's system peak. 4 pm.

A regression model was developed to estimate load reductions as a function of outdoor temperature and the activation of controls (MacDonald 1987). The net reduction depends strongly on outdoor temperatures (Fig. 15) because of two factors. First. the fraction of free riders (households participating in the program that do not have their air conditioners on at the time of control) decreases with increasing temperature. Second. the air conditioner duty cycle (percentage of time the units are on) increases with outdoor temperature. These results indicate that the program should be offered only to customers who use more than 600 kWh/month in the summer to reduce the incidence of free riders. and it should be dispatched only at temperatures above 100°F.

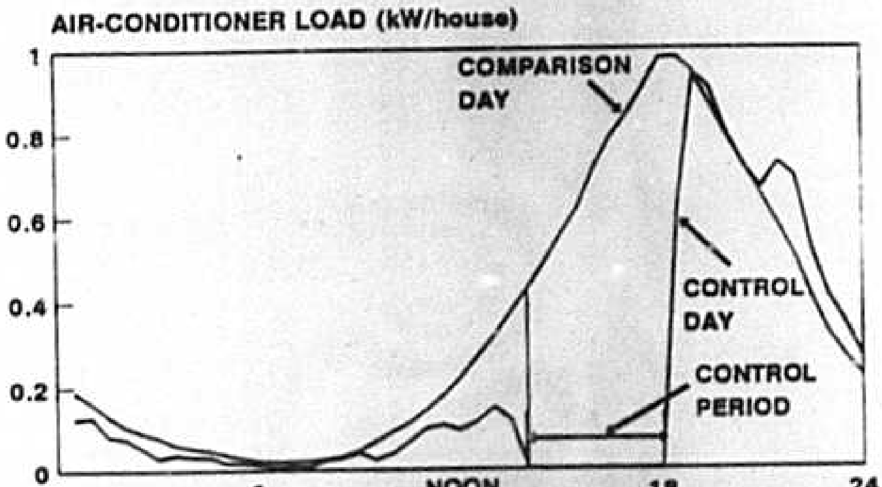


Fig. 14. Electricity demand for residential AC on control days and on noncontrol days. On control days, the units were turned off for four hours, from 2 to 6 pm.

Exhibit 9 (Continued)

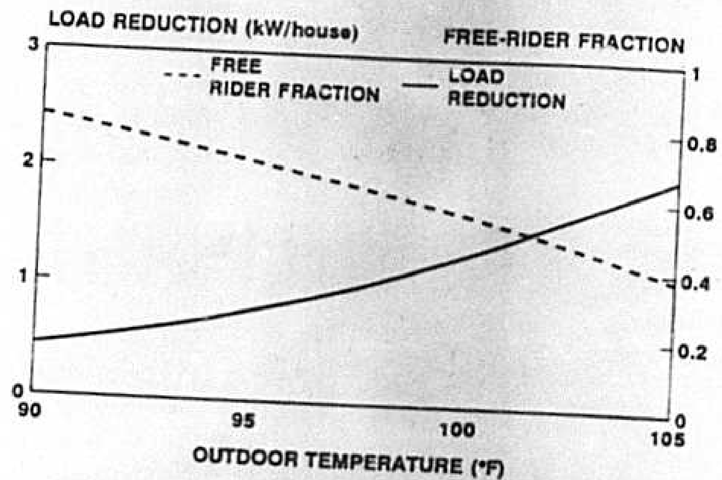


Fig. 15. The load relief achieved by PG&E's residential air-conditioner-control program increases with outdoor temperature.

devices strategically placed according to pump size and, perhaps, household size. However, the reliable measurement of central AC impacts requires more attention to statistical theory as well as recognition of the variations in customer response based upon such factors as geographic location, age of the head of the household, and age of the AC equipment.

Direct-installation programs (e.g., retrofits) and residential and C&I rebate programs (e.g., for light bulbs, refrigerators, or water heaters) have been used for a wide range of appliances and end-uses. The use of load research to measure these program impacts has been thought to be cost-prohibitive. Further, direct load research has limited applicability for assessing the impact of rebated measures.

Residential and C&I informational programs cover a wide spectrum of activities. They range from bill inserts informing customers about DSM programs to public appeals on radio and television to curtail load or cut back on energy use to customer audits that provide specific measure recommendations. Load research generally is not a viable measurement tool for information programs.

Residential and C&I time-of-use rate programs are excellent candidates for load-research-based evaluation. Load-research data are usually already available for C&I customers.

Load-curtailment programs have great potential to achieve significant peak reductions. Billing analysis is of little use in evaluating such programs unless it is accompanied by direct load-research data. While company-specific load research is usually preferable, borrowed data can be used in concert with billing analysis to estimate load-curtailment impacts in some instances. This information, coupled with weather and survey data, can be used to develop load-impact estimates for the customer, class, building type, and total.

JOINT UTILITY END-USE MONITORING

The decision to institute end-use monitoring to support DSM or other utility activities is a difficult one. Collecting end-use load data faces many barriers: cost, obtaining customer cooperation, the size of samples, and the many end-uses to monitor.

For these and other reasons, utilities sometimes hesitate to undertake end-use load monitoring on their own. As DSM programs become larger, some utilities will investigate alternative concepts for economically collecting end-use load data. One of these approaches is the joint utility project.

Joint utility projects can cut the costs of end-use monitoring. The major concern associated with end-use data analysis within a utility is the variability of the end-use loads. Adding multiple service territories may increase this variability, and joint projects must therefore be carefully considered. For example, weather, size of dwellings, type of equipment, construction practices, insulation, and family size can vary substantially within and among service territories, making SH a poor choice for joint end-use metering. Water heating may be a better prospect for joint projects because factors like family size, number of children, water source, and the presence of dishwashers and clothes washers can be identified and controlled. Estimates for the different service territories can be developed based on this information. The best prospects for joint end-use load research are those end-uses that are little affected by short-term weather and those end-uses that have little customer intervention (e.g., refrigerators, freezers, and water heaters) (Applied Energy Group, Inc. 1988).

CONCLUSIONS

Load research is a key element in the DSM-evaluation process. The costs of monitoring can be significant, but the need for measurement of DSM-program impacts makes this a necessary expenditure. This chapter identified a number of aspects of monitoring that the evaluator needs to attend to: planning, sample design, sample selection and validation, installation, quality control, and data analysis.

Direct monitoring has alternatives. Intelligent use of class, end-use, and borrowed load-research information, coupled with attribute information and statistical modelling, can often provide cost-effective alternatives to metering. The evaluator needs to carefully consider the problems and the alternatives before entering into a load-research program.

FREE RIDERS AND OTHER FACTORS THAT AFFECT NET PROGRAM IMPACTS

William Saxonis

INTRODUCTION

Accurately attributing energy savings and other benefits to DSM programs is a major goal of any comprehensive evaluation. To gain a better understanding of program impacts, the evaluation community has focused attention on program participants who would have taken the identical energy-conservation actions without the DSM program. This type of participant is referred to as a free rider.

Free riders must be examined because they represent a cost to the program but offer no direct benefits in return. A common practice among utilities is to adjust for free riders after gross capacity and energy savings are calculated. DSM programs that are highly successful from a gross-benefits perspective may prove to be less attractive when free riders are considered. For example, if 50% of participants are free riders, net program savings might be only half the gross savings. Program costs, however, would remain constant. From a societal perspective, free ridership is not important because they provide benefits to society even though their actions are not attributable to the DSM program.

In addition to helping determine net program impacts, free-rider data can serve as a useful tool in designing DSM programs for maximum effectiveness. By considering the impact of free-rider rates on program design, utilities could employ DSM resources more effectively and offer programs that are more responsive to the needs of the target audience.

While free ridership is theoretically an issue for all DSM programs, past research concentrated on programs offering rebates for the purchase of energy-efficient equipment and lighting. Appliance rebate programs have received considerable evaluation attention, resulting in at least 20 evaluation reports that include free-rider estimates. C&I rebate programs have also been frequent candidates for free-rider analysis.

To accurately identify the effects of free riders, understanding exactly what constitutes free ridership is important. Because programs can affect participants in different ways, three major free-rider categories are distinguished (Table 23).

A pure free rider is a participant who would have taken the *identical* energy-conservation actions at the same time if the program never existed. By contrast, a nonfree rider is a participant who takes the desired actions as a *direct* result of the program.

Table 23. Categories of free ridership

Category	Key characteristic
Pure free rider	Would have taken the identical action without the program
Incremental free rider	Influenced by the program to take action but not to the extent advocated by the program
Deferred free rider	Takes the action promoted by the program sooner than originally planned

An incremental free rider is a participant whose behavior is influenced by the program but not to the full extent advocated by the program. For example, a small business plans to install energy-efficient lighting without being aware of its utility's commercial-lighting program. After learning that the utility is offering a rebate for such lighting systems, the business purchases lighting that is 15% more energy efficient than the equipment it had originally planned to buy. In this case, the net program impact is the difference in electricity use between the lighting the business had planned to install and the even more efficient lighting it did install.

A deferred free rider is a participant who would have taken the same actions promoted by the program, but is influenced by the program to take the actions sooner than planned. For example, an evaluation of PG&E's Customized Rebate Program found that 68% of the participants planned to implement the same measures advocated by the program even if the utility did not offer a rebate. This group also said that the program encouraged them to install the measures from six months to more than five years sooner than originally planned (Pacific Consulting Group 1986). The energy savings between the time the customers installed the measures and the time they planned to take the actions are legitimately attributable to the program.

Unfortunately, defining free riders does not end with these three categories. For example, an incremental free rider may also be a deferred free rider.

FREE-RIDER MEASUREMENT

Despite the increased attention focused on free ridership, its measurement is an inexact science. Most of the literature on the topic has been critical of the reliability of the estimates. Lui and Fang (1990) concluded that "free-rider estimation can and should be improved." And Kreitler (1990) noted that "the standards for how to conduct evaluation research into free ridership need to be strengthened."

Although efforts are underway to improve measurement of free ridership, the lack of a reliable method to determine free-rider rates remains a problem. In fact, research has frequently produced confusing and inconclusive data (Table 24). For example, five utility C&I high-efficiency-motor programs identified free-rider rates ranging from 3% to 88% of total participants.

Table 24. **Examples of free-rider estimates**

Program type	Free-rider rates
Residential	
Refrigerator (rebate)	59–89%
Air conditioner (rebate)	19–79%
Heat pump (rebate)	40–60%
Loan programs	22–70%
Low-income programs	6–45%
C&I	
Lighting	5–85%
Motors	3–88%
Multiple end uses	5–70%

Sources: Kreitler (1990), Lui and Fang (1990), and Nadel (1990).

Several factors explain the variation in these estimates. In some cases, deficiencies in the research design cause the variance, but in other cases differences in program design, market segments, and reporting methods cause it (Table 25).

SURVEYS TO MEASURE FREE RIDERSHIP

Because programs and the environments in which they operate differ, the methods for determining the free-rider impact must be tailored to each program. Experience, however, provides valuable lessons for enhancing free-rider measurement with survey instruments, data from the marketplace, comparison groups, and modeling techniques.

Survey Program Participants

The most frequently used approach to measure free riders is to survey participants. Usually this survey asks participants one or more questions about what they would have done if the program were not available. This approach has a low cost, is easy to administer, and can provide results quickly. In some instances, such questions are part of the program-application form.

The key to successfully using this approach is to design carefully the questions aimed at determining free-rider rates. Kreitler (1990) claimed that a serious problem in accurately determining free-rider rates is that "inadequate attention has been devoted to the wording of survey questions, leading to misunderstanding on the part of respondents and misinterpretation on the part of analysts."

Table 25. **Different methods of reporting free-rider rates**

Wisconsin Electric refrigerator and air conditioner turn-in programs (single family)

Results
- 28% pure program influence
- 7% mixed program influence
- 28% mixed windfall[a]
- 37% pure windfall

Long Island Lighting residential air-conditioner rebate program (condenser-unit replacement only)

Energy Efficiency Ratio Range	Free-rider rate
10.5 - 10.99	10%
11.0 - 11.49	25%
11.5 - 11.99	22%
12.0 - over	15%

Northern Illinois Gas Company gas-furnace rebate program

Category	Free-rider rate
Pure free rider	71%
Incremental free rider	18%
Nonfree rider	4%
Do not know	7%

[a]Windfall reflects what would have occurred in the absence of the program and is equal to the sum of naturally occurring conservation and free ridership.

Sources: Wisconsin Electric (1991); Applied Energy Group (1989); and Lui and Fang (1989).

Target the Questions to Address Free Ridership

Some evaluations estimate free ridership by asking consumers to indicate if the program influenced them to take energy-conservation actions. While the responses may be useful in assessing the participant's feelings toward the program, they are a poor way to quantify the effects of free riders.

The principal flaw is that the question is too indirect to measure program-induced change. Participants may identify the program as being influential, but may also say that the level of influence was not sufficient to convince them to do something they would not have done anyway. This problem occurs in numerous evaluations. In an appliance rebate program operated by the New York State Energy Office, 71% of the participants claimed that the rebate influenced their appliance purchase, but only 34% indicated that they would have purchased a different appliance if the rebate were not available (Saxonis 1991).

In an evaluation of a commercial-sector program operated by Long Island Lighting Company, two questions were asked to determine free ridership. Participants were asked directly, "Would you have purchased that particular high-efficiency unit without the rebate offer?" Participants were also asked, "Did the availability of the rebate have any influence on your decision to purchase any of the technologies listed ... ?" In many cases consumers responded "Yes" to both questions. These responses are inconsistent. A sample of customers responding "Yes" to both questions was surveyed by phone to clarify their answers. The results of this survey found that almost all of the customers responding yes to both questions were free riders (Applied Energy Group 1989). Obviously, if only the second question was used to estimate free-rider rates, the results would be misleading.

Ask Questions That Provide a Complete Picture

Some evaluators simply ask participants if they would have taken the same action if the program were not available. For example:

- Would you have purchased this appliance if the rebate were not available?

- Would you have installed this measure without the program?

While this type of question directly asks about free ridership, a single question is usually not sufficient to fully address the issue. Asking questions similar to these will prompt only a yes (free rider) or no (nonfree rider) response. Incremental or deferred free riders cannot be identified with such simple questions.

An additional concern is that if the question is not well worded, the responses will be difficult to interpret (Kreitler 1991). For example, some consumers may confuse the purchase of the appliance in general with the purchase of a particular appliance model.

A better approach is to more thoroughly investigate the customer's decision to participate in the program. This approach involves asking several questions to validate responses and to gain a more complete understanding of participant behavior. In an evaluation of a C&I program operated by Madison Gas and Electric, customers were asked what they would have done without the program along with several questions to further probe and verify free-rider levels:

action before or after they heard about the program. Obviously, if they decided to take the action before hearing about the program, they are free riders or incremental free riders.

Sometimes, asking customers if they decided to take the action before becoming aware of the program can be challenging, especially if the DSM program involves several components (e.g., information, audit, and installation of measures). For example, a small business receives a utility-sponsored audit from an independent engineering firm. Later, the business talks to several contractors to learn more about the features and costs of several measures recommended in the audit, including some that are eligible for the utility rebate. Because of the various influences in this hypothetical program, it may be difficult for the participant to accurately identify and assess the impact of the DSM program as distinct from the independent influences of the engineering firm or the contractors. The customer might say that the utility program had no effect because a contractor convinced him/her to adopt the measure. In reality, the DSM program may have played a pivotal role in the firm's decision.

This problem was encountered in a C&I incentive program operated by Central Hudson Gas and Electric Corporation. The evaluation (Applied Management Sciences 1990a) states,

> A contractor or dealer who is aware of the program and intends to use the rebate as a marketing tool, might persuade the customer to invest in efficient equipment. When the customer fills out the marketing data on the application, however, it appears to him that he has intended to invest in efficient equipment from the beginning because that was the only type of equipment the dealer tried to sell him. In such a case, a self-reported free rider is not really a free rider.

Timing of the installation is also important in estimating free-rider rates. Did the program allow the consumer to take the action sooner than planned? If so, by how long? Table 26 shows sample questions for a C&I program operated by PG&E (Pacific Consulting Group 1986).

Identify Free Ridership for As Many Program Components As Possible

In programs that promote adoption of several DSM measures and practices, asking questions that pinpoint program elements where free ridership is the highest and where it is the lowest is important. This permits the utility to use evaluation results to fine tune the program by modifying or eliminating program measures that have high free-rider rates. A NEES program designed to encourage energy-efficient improvements in C&I structures illustrates this point. Rather than simply asking participants if they would have taken any actions without the program, respondents were asked to identify specific measures (e.g., energy-efficient lighting, heating-system improvements, and insulation) they would have implemented without the program. Although 53% of the participants indicated they would have installed some of the measures, the measure-specific free-rider rates ranged from 30 to 83%. This detail can be valuable in better understanding the program as well as in improving program design (Charles River Associates 1991).

Table 26. **Sample questions for a C&I program**

Question:	How did the availability of the 1984-85 Customized Rebate Program affect your decision to make the equipment change for which you got the rebate? Would you say ... ?
Responses:	You wouldn't have made the change without the program. You would have made the change anyway. You would have made the change anyway but the program speeded up the change. Don't know.
Question:	Did the 1984-85 Customized Rebate Program allow your company to undertake this rebate project *earlier* than originally planned?
Responses:	Yes No Don't know
Question:	(If Yes) How much earlier did you make the change?
Responses:	Less than 6 months 6 months to 1 year 2-3 years 4-5 years More than 5 years Don't know

Limitations of Survey Methods

Even with well-designed survey questions, the evaluator must deal with recall and other problems of self-reported data. Specific issues include fading memories, a tendency to tell surveyors what they want to hear, and a reluctance to admit that they would have done something different from what they did.

Questions on the energy-related actions of participants may be asked several months after their involvement in a program. It may be unrealistic to expect participants to remember all the details of their participation. As a result, evaluators should survey participants as soon as possible after they receive program services.

Some respondents are reluctant to admit taking advantage of a program that failed to influence them. Others may not know what they would have done without the program: they may provide answers that they think will please the interviewer and will be consistent with positive goals, such as protecting the environment.

Do people learn, believe, and then act, or do they act and then make their beliefs consistent with their actions? Once a purchase is made, the consumer's perception of the product not selected may become more negative, and the perception of the product purchased may become more positive. This phenomenon makes it unlikely that the consumer will indicate a desire to have taken a different action without the program service (Calder 1973). Such responses will result in an upward bias in free-rider estimates.

MARKETPLACE DATA TO ESTIMATE FREE RIDERS

Surveying program participants is not the only way of determining free-rider rates. Other methods include analysis of market data, use of comparison groups, and discrete-choice modeling. These topics are discussed in this and the following two sections.

The impact of DSM programs is usually not limited to the customer that actually receives the service. If consumers are receiving rebates for purchasing efficient appliances, retailers, distributors, and manufacturers may act to ensure that the qualifying products are available in the stores. Moreover, they may promote the program by using the utility incentive as part of their own marketing efforts. For programs that advocate measures that require professional installation, a network of architects, engineers, contractors, installers, and suppliers are affected. These "trade allies" (service providers, equipment distributors, and retailers) can be surveyed to yield estimates of free ridership independent of those obtained from program participants.

For appliance-rebate programs, the net impact of the program appears to be higher (and therefore estimates of free ridership are lower) when sales data (rather than participant surveys) are used to estimate free riders. Dealers participating in the New York State Appliance Rebate Program reported a significant shift in the number of energy-efficient refrigerators and room air conditioners they stocked and sold during the time the rebates were available. This occurred at a time when only a modest increase occurred in the efficiency levels of appliances sold nationally. A majority of the dealers attributed the increase in efficiency levels directly to the program. Anecdotal reports suggest that appliance distributors shipped less-efficient appliances to other regions (Saxonis 1989). Considering that the primary purpose of the program was to increase the sale of efficient appliances, the program may have achieved its goal despite an apparently high free-rider rate (as estimated from participant self-reports). The appliance dealers appeared to have successfully reduced the opportunity for consumers to purchase inefficient models.

In the case of rebate programs, evaluators have asked retailers if they sold more of the product qualifying for the incentive during the program than before the program. After adjusting for general trends in product sales, the difference between the before and after sales provides an indication of the net impact of the program. This approach has the advantage of examining what actually occurs as contrasted to asking consumers to speculate on what actions they would have taken if the program never existed.

A major barrier to collecting this type of information is the reluctance of retailers to provide data because such information is often considered proprietary. National trade organizations, such as the Association of Home Appliance Manufacturers and Gas Appliance Manufacturers Association, provide some product and sales statistics, but they usually reflect only national trends and fail to identify local sales of specific models.

Despite these difficulties, several approaches have been used successfully to collect marketplace data. A key to success is developing a positive relationship with the organizations from which you are requesting the information and stressing the importance of the evaluation effort to their business. Wisconsin Electric reported that, with persistence (i.e., several phone calls and a letter from a senior vice president), they received sales data from an initially reluctant group of lighting distributors. They also encouraged cooperation by paying the distributors $200 each to help defray the cost of collecting and reporting the data (Brooker and Fichtner 1991). Wisconsin Electric also obtained sales data by paying appliance dealers $5 per sale to collect data on purchases of both qualifying and nonqualifying air conditioners as part of an evaluation of their Smart Money program (Brugger and Brooker 1991).

Some utilities require dealers to provide sales data as a prerequisite for participating in their program. For example, the San Diego Gas & Electric Earthwise Appliance Program requires participating dealers to provide both preprogram and program sales data. Most of the major-appliance dealers in the San Diego area (about 60) have agreed to participate. Some dealers have even offered to provide names and addresses of consumers who purchased appliances not eligible for the rebate. In addition, the utility plans to conduct a survey of appliances displayed in the dealers' showrooms for additional data validation (Wiggins 1991).

Another approach is to survey dealers with a voluntary and confidential survey. An evaluation of an appliance rebate program in New York asked dealers to indicate in ranges (e.g., 10-20%, 30-40%) the percentage increase or decrease in inventory of equipment that would qualify for the rebate. Ranges were used to reduce dealer reluctance to provide detailed sales and inventory data. Dealers were also asked if inventory changes were made because of the program (New York State Energy Office 1989).

The effectiveness of this type of data collection is limited because ranges sacrifice precision. For example, if the dealer indicated that sales increased 10 to 20%, the actual percentage could be as low as 10% or as a high as 20%.

An additional concern is that dealers may overestimate program impacts to encourage continuation or expansion of the program. It is important to consider that the availability of the DSM program may benefit their business.

Nonresponse bias is a potential problem for all of these approaches, if not all the affected dealers participate. It may be possible that the dealers most enthusiastic about the program will respond to the survey and the less enthusiastic dealers will not.

COMPARISON GROUPS TO ESTIMATE FREE RIDERS

The use of a comparison group is a common evaluation technique that can be employed to account for free riders. The difference between the energy-conservation actions of a sample of program participants and an appropriately selected comparison group should provide an accurate estimate of the net program impacts. A major advantage of this approach is that it compares actions as opposed to asking participants to estimate what they would have done if the program were not available.

Finding a group that is comparable to the program participants but not contaminated by the DSM program can be difficult. Specifically, participants choose to participate while nonparticipants either choose not to participate or are unaware of the program. As a result, systematic differences may exist between participants and nonparticipants with respect to energy-use patterns and attitudes towards energy efficiency.

Addressing the potential for self-selection bias in the comparison group is important. This point was vividly illustrated in a process evaluation of NEES's C&I incentive program. The evaluators attempted to estimate the level of free ridership by comparing the activity of participants to nonparticipants. They found that 13% of the nonparticipants installed measures similar to those being promoted by the program, suggesting a free-rider rate of the same percentage.

The report correctly noted that differences between participants and nonparticipants would likely inject a significant bias in their free-rider estimate. Specifically, "participants appear to be different from nonparticipants in several interesting ways. Participants tend to look to outside sources of expertise for advice while nonparticipants are do-it-yourselfers." This finding suggests that free-ridership estimates that are based on comparisons to nonparticipant actions may be flawed, because participants may need the program to undertake the retrofit measures, while nonparticipants may not (Freeman Research Resources 1991).

In some cases, the differences between participants and nonparticipants will be less dramatic and will not seriously affect results. If differences are significant, evaluators can correct for self-selection bias. Methods for doing so include comparing energy consumption between participants and nonparticipants using multiple comparison groups and simultaneous-equation models. The models generally require the estimation of two equations, a discrete-choice participation model and an energy-use equation (Chapter 4).

Another approach is to monitor the activity of a nonparticipating region with similar characteristics to the test market. Wisconsin Electric used this approach to evaluate a program to encourage C&I customers to purchase energy-efficient fluorescent lamps and ballasts. Equipment sales were monitored in the area where the rebate was available (Milwaukee) and in a region with similar characteristics (e.g., population and electric rates) but without a lighting-efficiency program (Fig. 16). Cincinnati was selected as the comparison region. Sales of efficient lighting fixtures remained nearly constant in Cincinnati, suggesting that the increase in the sale of efficient lighting in Milwaukee during the program was directly attributable to the program (Brooker and Fichtner 1991).

Deferred free riders should be considered when comparison groups are used (as well as with other methods). In a comparison-group analysis, the effects of deferred free riders would appear as a decrease in net (program-induced) energy savings over time; see Chapter 7.

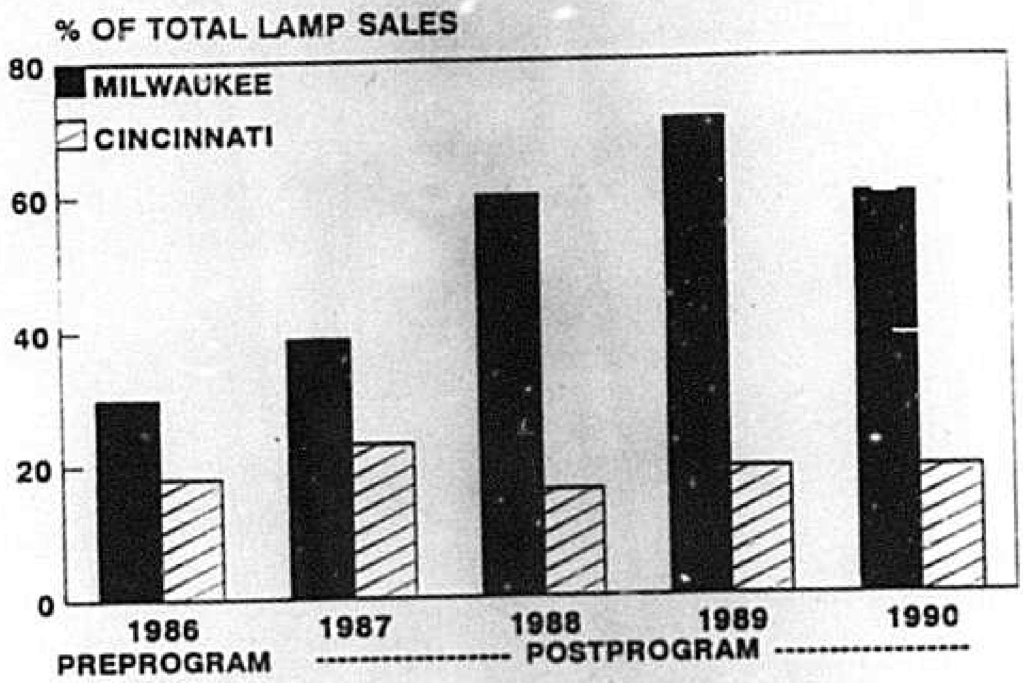


Fig. 16. Comparison of sales of energy-efficient lamps in Milwaukee, Wisconsin, where a program operated and Cincinnati, Ohio, where no such program operated.

STATISTICAL MODELS TO ESTIMATE FREE RIDERS

Some evaluators have used statistical models to estimate free ridership (Chapter 4). Generally, this method involves analysis of the energy-related actions, characteristics, and attitudes for samples of participants and similar nonparticipants. Simulations are developed to predict the likelihood of the adoption of program-sponsored measures with and without the program. The two estimates are then used to calculate the free-rider ratio.

Niagara Mohawk used this approach to estimate the free-rider rate in its Low Cost Measures Program (Regional Economic Research 1991). This program offered low-flow showerheads, pipe wrap, water-heater blankets, and compact fluorescent bulbs to residential customers. With data from a survey of participants and nonparticipants, an equation was developed to estimate the likelihood of the adoption of these measures. Characteristics like income, household size, education level, and attitudes toward conservation were used to adjust for differences in participants and nonparticipants (self-selection bias).

The estimated model (Table 27) was used to simulate the predicted adoption rates of program participants under two scenarios: customers participate in the program and they do not. This dual simulation involves solving the model for the participant sample with different values for the participation variable (yes vs no).

Table 27. **Example of statistical model used to estimate free ridership in a residential program**

$$ADOPT_{ik} = A_k(OWN_i, INC_i, NUMINHH_i, EDUC_i, PART_i, ATT_{1i}, ..., ATT_{Ji}, MILLSRAT_i),$$

where:

$ADOPT_{ik}$ = a binary variable reflecting the adoption of measure k by household i

OWN_i = a binary variable reflecting home ownership

INC_i = income of household i

$NUMINHH_i$ = household size

$EDUC_i$ = education level

ATT_{ji} = an attitude index (j) relating to conservation

$PART_i$ = participation of the household in the program

$MILLSRAT_i$ = Mills ratio (used to correct for self-selection)

Because the dependent variable is binary, a logit formulation was used, which assumes the logistic functional form:

$$ADOPT_{ik} = 1/(1 + e^{-h_k(x)}),$$

where: $h_k(x) = h_k(OWN_i, INC_i, NUMINHH_i, EDUC_i, PART_i, ATT_{1i}, ..., ATT_{Ji}, MILLSRAT_i)$

Source: Regional Economic Research, Inc. (1991)

The specifics of the model will vary with program type and availability of data. For example, in a residential loan program, factors like the number of people over age 75, the number of people at home at night, and the importance of comfort were considered (Ozog 1991). In modeling a commercial-sector program, variables would likely include the type of business and its energy-consumption history (Regional Economic Research 1990).

While these techniques hold promise, they are hindered by the quality of the data on which they are based. These data usually suffer from the same problems as other evaluation techniques. Regional Economic Research (1991) noted, "The adoptions model was estimated using the results of the tracking survey. While the survey yields a relatively small sample size, it is the only available consistent source of information on adoptions and nonparticipants. In the future, with more data points, our free-rider estimates may change."

FREE DRIVERS

The opposite of a free rider is a free driver. A free driver contributes to the goals of the program (e.g., reduce energy consumption) but is not formally a program participant. A free driver is affected by the program either through a conscious awareness of the program or because of program-induced changes in the marketplace. Free drivers require evaluators that use comparison groups to consider whether the comparison group is actually taking the conservation actions because of the program (i.e., what is traditionally considered a free rider may turn out to be a free driver); see Chapter 7.

An example of a free driver is a customer who purchases a product that qualifies for a rebate but does not claim the rebate. Northern States Power Company's analysis of its appliance rebate program showed that only about 40% of the customers that purchased a qualifying appliance applied for the rebate (Brian Gard William Lesh Inc. 1986). In some cases the customers may not have been aware of the program, and in other cases may have decided it was not worth the effort to apply for the rebate.

Free drivers also occur in new-construction programs. Research conducted on such programs in Maine and Wisconsin found that they affected the actions of nonparticipants. Specifically, some builders constructed homes to program standards for competitive reasons but chose not to participate in the program. As a result, an examination of nonparticipants would not accurately represent what would have happened if the program did not exist (Violette, Ozog, and Wear 1991).

Research on free drivers is limited. One way to identify free drivers is to ask nonparticipants if they are aware of the utility's DSM program and, if so, whether the program influenced their adoption of program-recommended actions. Unfortunately, customers may be influenced by a program even if they are unaware of its existence. For example, wholesalers and retailers may stock more of the energy-efficient appliances because of a utility's rebate program: nonparticipants may purchase these units even if they are unaware of the program simply because the efficient ones are in greater abundance and more prominently displayed in the showroom.

Another approach is to use comparison groups consisting of communities outside the area in which the program is offered. Comparing the distributions of efficiencies for program-sponsored measures in the participating area with those in the comparison area might show the overall effect of the program, including (implicitly) the contributions from free drivers.

Another approach is to survey the trade allies involved in a program. Their responses to program-induced changes, both at the distributor level and the customer level, might identify the degree of free drivership in a particular program. For example, a retailer not participating in a program might stock the more efficient units anyway to remain competitive with the participating retailers, an example of free drivership among trade allies.

DISCUSSION AND CONCLUSIONS

It may not always be necessary to conduct rigorous analysis of free ridership because the importance of free riders varies with program type. Free-rider measurement is usually more important in determining program cost-effectiveness where the benefit/cost ratio is close to one.

An appliance rebate program offering generous rebates for the purchase of energy-efficient products would be expected to be sensitive to free-rider rates in determining net program impacts. Not surprisingly, most of the free-rider research has been targeted toward this type of program. On the other hand, a low-cost energy-outreach program designed to encourage people to insulate their attics could tolerate a high level of free ridership because the cost of the program is low and the benefits are high. Even more costly programs, such as those offering energy audits, may be able to tolerate free-rider rates of 50% or more and remain cost-effective. In BPA's industrial programs, program designers were generally not concerned about free riders because, according to Keating (1990):

> The savings are so inexpensive that the cost-effectiveness limit wouldn't even be approached at 50% free riders and there was a clear possibility that without the utility's intervention and quality control, the measures that might have been installed would not have been as comprehensive or aggressive. [In addition] the utility did not want to risk the loss of such large blocks of savings (1-8,000,000 kWh/yr) if the evaluators were wrong about the willingness of the participants to install the measures without the program.

DSM programs aimed at low-income customers have experienced low free-rider rates. Research suggests that low-income households don't have the money to implement major conservation measures without assistance. Most studies of such programs do not incorporate a free-rider test, and those that have often found rates below 15% (Lui and Fang 1989).

Free-rider levels may also be unimportant if a program is designed to move a specific market (e.g., encourage retailers to stock energy-efficient light bulbs and high-efficiency appliances). If this goal is achieved, the free-rider rate is irrelevant.

One way of dealing with free riders is to design programs to reduce their impact. Free-rider research has been valuable in this regard. For example, in a commercial lighting program offered by NEES, the free-rider rate was quite high, about 65%. Rather than eliminating the program, the program was modified. The first change was to require preinstallation inspections to ensure that potential participants did not already have efficient lamps and ballasts and merely wanted to replace worn-out equipment. This type of customer would readily fit the description of a free rider. The utility also encouraged the use of "advanced" lighting measures that typically have free-rider rates considerably lower than simple conversions from incandescent to fluorescent fixtures. The free-rider rate for regular fluorescent lamps was 65%, but for compact fluorescent lamps only 5%. In general, free-rider rates tend to be higher for products with high market shares and lower for products with low market shares (Nadel 1990).

Research conducted on a rebate program for C&I customers operated by Southern California Edison shows how different segments of the target market reacted to the program. The program gave rebates for the installation of energy-saving equipment. The customers who asked to be audited (as opposed to a utility representative's initiating the audit) had a greater tendency to install recommended measures (Train, Ignelzi, and Kumm 1985). These results show that free ridership is a function not just of program design but also of the type of customer participating in the program. Evaluation results can be used to help utilities better target their programs to customer groups with low free-rider rates.

Markets, customer attitudes, and actions are not constant. A free-rider rate of 10% today may be 50% tomorrow. Despite the problems associated with free-rider measurement, if questions are asked over a long time with the same measurement techniques, important trends can be detected even though the absolute accuracy of the estimate may be questioned.

Measuring and dealing with the free-rider issue is not simple. The data can be inconclusive and contradictory, which makes it risky to depend on a single method of collecting and analyzing free-rider data. The challenge for the evaluation community is to improve techniques for free-rider measurement. The challenge for those developing DSM programs is to reduce the free-rider impact in the design stage.

PROCESS EVALUATION OF DSM PROGRAMS

Benson Bronfman
Jane Peters

BACKGROUND

Process evaluation is the review and assessment of program implementation. Originally applied to federal housing and education programs, process evaluation plays a central role in the evaluation of DSM programs. The major characteristic of DSM programs that distinguishes them from power plants is the distributed nature of DSM-program activities and effects. DSM programs are centrally planned, but are implemented by field staff, contractors, and customers.

Initially, the focus of process evaluation was on program operations and structure, and such evaluations provided program managers with a documented history of the program. Currently, the goal of process evaluation is program optimization through

- Improvement in implementation efficiency
- Assessment of market segments and targeting of specific segments
- Improvement in quality of measure installation
- Identification of program-design issues
- Interim accounting of program progress through reviews of the program database
- Examination of special issues, such as measure life and program comprehensiveness

Despite the creative uses of process evaluation, too often process evaluations are narrowly defined to include only customer response and satisfaction. While satisfaction is clearly important, programs could achieve high satisfaction but acquire few DSM resources. Measuring customer satisfaction, in and of itself, does not represent the current practice of process evaluation.

Evaluation includes a continuum of activities designed to measure program impacts and to improve program delivery. Because of this range, the attribution of activities exclusively to "impact" or "process" practice is somewhat arbitrary. Impact evaluations often include surveys of consumer satisfaction and behavior. Process evaluations frequently document program activities (e.g., measures installed and their actual costs) and estimate cost-effectiveness. This review of process evaluation takes a broad view of the field, including some activities that might belong on the impact side of the ledger.

As DSM-program performance has become more closely tied to integrated resource planning and incentive regulation, process evaluation has assumed more importance. First, because many utilities have only recently come to the DSM-resource-acquisition philosophy, many programs are designed and implemented quickly. Even with expert assistance in program design, early process evaluations are critical to fine-tune programs, assess the market potential, and identify market segments not reached. Second, because process evaluations are in the field early and because impact evaluations often do not produce results for a year or two, process evaluations are instrumental in producing early reports on program effects and effectiveness. This activity may include assistance in designing the program database, assessing the reliability of program data, and periodic reporting of implementation data (e.g., the number of installations, estimated and actual costs, and projected savings).

Process evaluations are generally conducted once a year for each program. The topics to be addressed vary from year to year, depending on the stage of the program and the types of issues that are currently important to the program. Start-up evaluations often address program design, contract-negotiation procedures, and internal organization. Short-term evaluations can be conducted to address special issues, such as changes in building codes, the effects of utility reorganization, or changes in the marketplace. Comprehensive evaluations may include a variety of projects designed to address specific problems from start-up to special issues, finally combining all the results into a comprehensive report on the program.

Process evaluations should be conducted with guarantees of confidentiality to all parties providing information. This guarantee will help demonstrate that the evaluation will be used to improve the program. It also serves to assure customers that the utility is concerned with the program as a whole and not with their particular response to the program. By assuring all contacts that the information is confidential, greater objectivity and more insightful reports will be obtained.

Process evaluations can be conducted by utility staff or by contractors. Utilities have often conducted their own process evaluations because of the sensitive nature of the information and because process evaluations are primarily for internal use. However, the use of contractors is becoming more accepted because contractors can provide an alternative and confidential channel for communication among program staff, participants, contractors, and utility decision makers. An additional benefit of contractors often occurs after a utility reorganization. Contractors frequently serve as the corporate memory during these changes.

KEY ISSUES

Process evaluations are broadly defined as the assessment of program implementation. A comprehensive process evaluation examines a program at different stages. This section discusses the planning and design stage as well as the implementation stage (Exhibit 10).

Exhibit 10. Process evaluation of a DSM-bidding program

In 1988, Orange & Rockland Utilities proposed to run an experimental bidding program for DSM resources. The New York State Energy R&D Authority assisted in the design of the program and sponsored an evaluation of the program to inform the other utilities in New York (Environmental and Energy Services Company and Pacific Energy Associates 1989; Peters, Barry, Horowitz, and Gordon 1990).

The evaluation included two phases. The first addressed program planning and design, contractor selection, and contract negotiations. The second addressed the first year of implementation, including an analysis of project savings and costs. The evaluation addresses seven major issues:

1. Program Goals and Design. Program goals provide a framework to determine whether and how policy intentions are framed within the program design; whether the program is being implemented as intended; how the goals are perceived by contractors; and, ultimately, whether the program will be manageable as it is implemented.

2. Request for Proposals (RFP) Process. With surveys, the bidding process was reviewed to determine the level of competitiveness and whether program goals were achieved in the selection process. The reasons for nonresponse were used to determine if the utility had an appropriate list of potential bidders and whether any features of the program or RFP presented barriers to response.

3. Delivery Mechanism. The major issues to assess were the costs borne by the contractors and customers, the relationship between the utility and the contractors, the relationship between the contractors and utility customers, and the ability of program contractors to deliver the DSM savings.

4. Administrative Mechanism. The administrative components of the program included development of the RFP, the contracting process between the utility and the ESCOs, the administrative costs borne by different parties, the cost- and quality-control procedures used by the utility, and the savings-verification process.

5. Customer Response. Customer response and satisfaction are key ingredients to program success. However, the utility was not a direct party to the customer's experience because the program was delivered by ESCOs. The evaluation obtained the customers' view of the utility's role in the program, their response to marketing materials, and their view of the ESCO role. In addition, the evaluation addresses the effects of franchising on customers, contractors, and the utility.

6. Implementation. This phase estimated energy and demand savings and program costs. These data were then used to assess cost-effectiveness.

7. Measurement Plans. The validity of the cost-effectiveness analysis rests on the accuracy of the cost and savings data for the installed measures. The cost data were obtained during the evaluation of the administrative and delivery mechanisms. Collecting actual savings data requires detailed estimates of the impact of the installed measures on consumption and demand. Each contractor was required to develop a measurement plan. This phase of the evaluation reviewed the plans and determined if additional data were required to assure the validity and reliability of the savings data.

Planning and Design

Process evaluations can address program-design issues before the program is in the field. Evaluations conducted during this stage can lead to early program enhancement and more effective long-term implementation (Peters, Gustafson, and Vowles 1987; Peters, Barry, Horowitz, and Gordon 1990). The most critical issue for investigation during this period is to determine whether the goals of the program are explicit and whether they are being met. Lack of clear articulation of goals can have severe implications for future evaluations of program performance (Lerman and Bronfman 1986a). Process evaluations are designed to determine if programs are achieving their goals. Frequently, program goals change over time. These changes should be documented, and program design should be reviewed to determine whether it supports the new goals.

Assessing the planning and design process requires early review and feedback. The evaluator seeks to determine whether the design process successfully addressed all parties' concerns and whether the resulting program has sufficient staff resources and commitment for effective implementation. Experience at other utilities with similar programs is usually brought to bear to review the proposed program structure, delivery mechanism, and administration.

Implementation

Program implementation includes administration, program-delivery mechanism, ongoing program activities, and customer response to a program. Most process evaluations focus on this phase, which can be repeated at annual intervals to track the program over time. For example, Lerman and Bronfman (1984); Lerman, Bronfman, and Tonn (1983) and Synergic Resources Corporation (1991) conducted process evaluations of BPA's Residential Weatherization Program as it developed from a pilot to an interim to a long-term program. In another example, the early adoption of model conservation standards in Tacoma, Washington, led to the implementation of a kick-off evaluation as the program was starting, and a revisit to examine implementation as the program matured (Lerman and Bronfman 1986b). Several process evaluations that tracked the evolution of an industrial retrofit program are chronicled in Evans and Peters (1989).

Administration mechanisms range from contractual agreements between the utility and customers or contractors, to payment procedures and quality- and cost-control procedures. Issues about contractual agreements are only addressed when contracts must be signed, while payment and quality- and cost-control procedures should be monitored throughout the program to assure that they are functioning well.

Customer response to a program determines whether it is effective or not. Customer response includes such issues as free riders, take back, free drivers, market effects, and market segmentation. If customers are dissatisfied with the program or if they find participation difficult, the program may require redesign.

How the program is organized and how the utility manages a program are critical aspects of program process. For example, if the utility permits districts to implement the program in their own way, several programs will be in the field, and a variety of explanations will exist for program effects (Fig. 17) (Lerman, Bronfman, and Tonn 1983). Reviews of how utility staff communicate within a program can reveal interesting information about implementation. For example, Lerman and Bronfman (1984) showed that the delay in resolving technical issues and communicating those results to field staff led to ad hoc decision making in the field and infrequent use of the required technical-issue-resolution process.

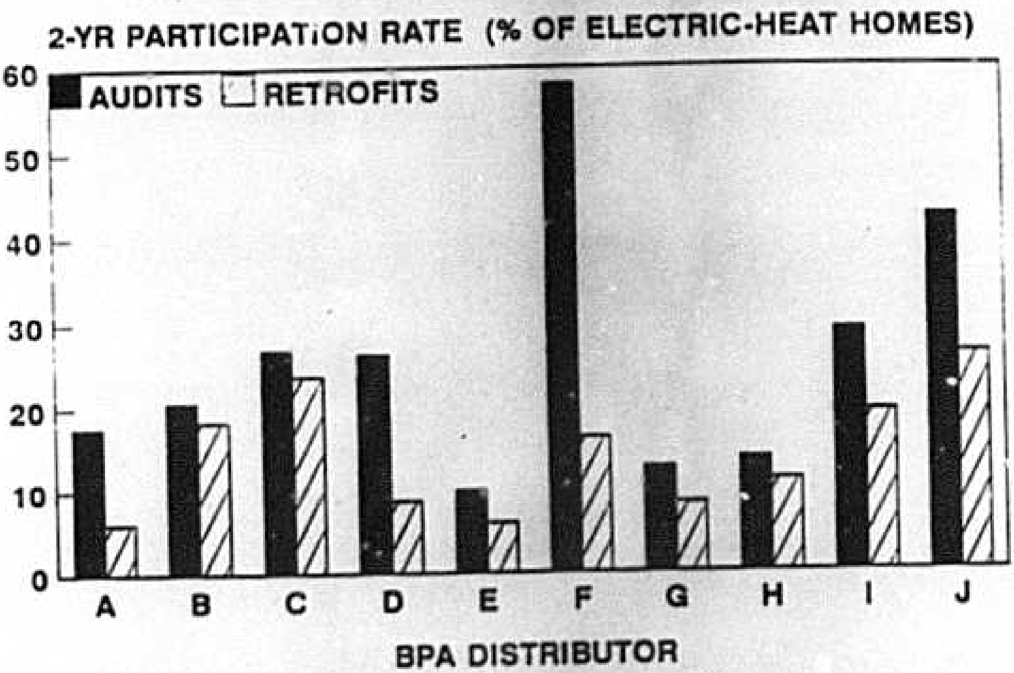


Fig. 17. Customer participation rates across the ten utilities participating in the BPA pilot Residential Weatherization Program. Because these utilities each implemented the program differently, customer response to the energy audit and financing of retrofits differed substantially.

A critical question for process evaluations is, "How much of the target market participated in the program?" Market segmentation helps to identify appropriate groups of customers to target for a particular program. The most common approach to market segmentation is to link attitudinal and demographic information obtained from surveys or geographically with specialized databases (Van Liere and McKinnell 1989; Feldman 1988). Process evaluations seek to determine whether the expected segments participated in the program, whether the segmentation strategy was effective, and how it can be improved. For example, one evaluation found that the structure of investment decision making among manufacturers and their place in the annual budget cycle was much more important in determining program participation than was company size or type.

Barriers to participation are the factors that prevent customers from participating in a program. A difficult sign-up process or confusion about program requirements or benefits can limit participation. Two evaluations examined the low response rates to RFPs for a third-party financing commercial program. These evaluations showed that the low response rates were, in large part, caused by the many firms receiving the RFP who were either not qualified to respond or had no intention to respond.

Process evaluations address nonparticipating customers as well as participating customers to identify the program features that are least attractive or that require modification to improve participation rates. Two evaluations examined the barriers to participation in commercial retrofit and new-construction programs (Freeman Research Resources 1991; Charles Rivers Associates 1991). Both studies linked reasons for participation and nonparticipation to specific program features.

Utilities often want to estimate the number of free riders (Chapter 8). Identification of free riders is often accomplished through surveys of participants and nonparticipants (Charles Rivers Associates 1991; Freeman Research Services 1991).

Customers must be aware of the program before they can participate. Measurement of program awareness focuses on nonparticipants in the target market. Awareness can mean that they have heard of the program or that they are aware the utility conducts energy-efficiency programs in general. Awareness following media campaigns can be expected to rise. Determining how much rise occurs after different types of campaigns can help utilities select the most appropriate strategies for future marketing efforts (Birnbaum and Davis 1989; Columbia Information Systems 1988).

Customer satisfaction includes general satisfaction with the utility and satisfaction with specific programs. Changes in customer satisfaction can be measured by repeated studies over time. Satisfaction with DSM programs is critical for understanding the benefits of the program and making modifications to improve customer response. If the utility conducts general-satisfaction studies as well, participants can be compared to nonparticipants (Peters, Haeri, and Seratt 1990).

Utilities frequently examine attitudes and beliefs to determine how they affect participation and how participation affects attitudes. For instance, it was found that attitudes toward the utility can be improved and customers' sense of control over energy use can be increased through participation in residential low-income programs.

During the lifetime of a program, the particular technologies promoted by the program can advance beyond the levels included in the program, or the energy-efficient technology can become the dominant option offered in the marketplace. Measuring these changes is an important part of process evaluations. Studies of retailer and distributor inventories, mystery-shopper activities, and discussions with equipment vendors and dealers can help determine whether a program has outlived its usefulness or whether the installation of new equipment should be encouraged through changes in program design. Two process evaluations showed, for example, that more than 75% of eligible street lighting had been replaced in BPA's Street and Area Lighting Program (Peters and Bronfman 1986) and that the high efficiency water heaters were the only options available in Washington and Oregon (Lerman 1987).

Program costs include incentives, utility staff, and contractor expenses. In addition, customers may have costs associated with the program. Surveys can be used to identify costs that are not recorded and to estimate costs. For example, Peters, Haeri and Gustafson (1988) estimated administrative costs for three commercial programs with a combination of utility records, surveys, and in-person interviews.

DATA COLLECTION AND ANALYSIS

Process evaluation relies on traditional social-science methods to collect and analyze data. Data sources include utility staff, trade allies, program participants, nonparticipants, implementation contractors, program reports, program databases, and data from other utilities conducting similar programs. Data-collection methods include structured in-person interviews, group interviews (e.g., focus groups), mail and telephone surveys, direct observation, and literature and document reviews. Analytic techniques range from narrative description and anecdotal reporting to simple and multivariate statistical analyses.

Data

Identification of the appropriate data sources requires a sound understanding of the program's design. The evaluation staff must identify the key people involved in the program within the utility. The following questions (see also Exhibit 11) can be used to identify these key people.

- Who designed the program?
- Who approved the program design?
- Who delivers the program to customers?
- Who administers contracts and rebates?
- Who interprets questions about the program?

Once the correct contacts are identified, program documents and communications among these people should be located. Utility contacts can assist evaluation staff to identify the role of contractors and trade allies in implementing the program. The program database and monitoring system provide additional information.

A process evaluation of an all-source bidding program identified six target populations (Environmental and Energy Services Company and Pacific Energy Associates 1990):

Exhibit 11. Key objectives of process evaluations for Boston Edison Company

Boston Edison (1991) prepared a series of questions to address in evaluating each of its DSM programs. The key objectives for evaluation of a residential program are:

- Document the history and progress of the program
- Assess the promotion and delivery system for use in designing future programs
- Assess installation-contractor effectiveness and comprehensiveness in program delivery
- Assess building-owner and occupant satisfaction and attitudes toward the program and its products and services
- Identify barriers to program penetration
- Assess the market-driven versus program-motivated efficiency actions of customers to estimate the numbers of free drivers and free riders
- Evaluate product removal and retention to estimate persistence of savings
- Identify and compare the characteristics of owner and occupant participants and nonparticipants
- Assess the remaining market potential
- Examine the roles played by contractors, community organizations, building owners, occupants, and Boston Edison staff in the design, marketing, and implementation of the program
- Assess the educational components of the program and its effect on participating customers

- Regulators in New York and New Jersey involved in planning the program
- Utility staff involved in the design, development, and implementation of the program
- ESCOs and utility customers submitting proposals and signing contracts to deliver DSM resources
- Nonresponding ESCOs and customers (those on the RFP mailing list who did not submit a proposal)
- Participating customers who received program services from ESCOs
- Nonparticipating customers who considered participation but then decided against it

Selection of the appropriate number of contacts and samples for surveys requires understanding the difference between sampling for information and sampling for statistical reliability. Sampling for information provides a wide enough range of experience with the program to cover all points of view. This is the approach generally taken with key-contact sampling for the utility staff, implementation contractors, and small numbers of participants.

Sampling for statistical reliability assures that statistical analyses yield the desired precision in results (Chapters 3 and 4). Statistical reliability is especially important in impact evaluations. Samples should be constructed so that the results can be generalized to the population as a whole. While general guidelines suggest that a sample of 200 to 500 is sufficient in most cases, some situations require larger samples to ensure that every market segment is included in the analysis.

If the program is new to the utility, drawing on the experience of other utilities to identify issues and approaches for solving problems is appropriate. Literature reviews and surveys of other utilities' experiences can also be valuable when addressing difficult issues, such as free ridership or measure persistence.

Process evaluations can occur throughout the life of a program, including the first few months of program implementation. Such evaluations are used to obtain an early assessment of the program process. They are conducted over a short time and rely on interviews as the primary data-collection method.

Process evaluations conducted during the first few years of the program are generally comprehensive. These evaluations focus on both planning and implementation, and are often coordinated with a preliminary impact evaluation. Such evaluations require extensive interviews and surveys with many people involved in the program.

Evaluations that occur in the middle and later periods of the program might focus on specific issues of implementation or whether specific recommendations have been adopted. Process evaluations might collect data required for the impact evaluation or to determine if the market has changed so that the program is no longer required. These special evaluations require targeted interviews and surveys of appropriate populations.

The final stage of a program may require no evaluation at all. If only limited studies have been conducted, however, a final, comprehensive evaluation may be undertaken. Such an evaluation can help interpret impact data and provide a resource for utility staff who might wish to build upon the experience of the program in the future. These evaluations would also require extensive interviews with many contacts and surveys of large samples of participants and nonparticipants.

The depth of the analysis required will also influence the selection of a data-collection approach. In general, the greatest depth is gained from in-person interviews, which can provide detailed and candid accounts of the program process. Surveys, focus groups, and document reviews provide less detail but permit access to a greater number of contacts. In programs with many participants, depth must generally be discounted to achieve breadth of coverage for participant and nonparticipant views.

The number of contacts is an important factor in determining the breadth of coverage required in collecting data. If there are many potential contacts, those with unique and comprehensive familiarity with the program should be identified. These contacts will be the most appropriate for in-person interviews. Field staff or trade allies with comprehensive implementation experience can be effectively interviewed in groups. Surveys are most appropriate when breadth is required, such as with large groups of participants, nonparticipants, and trade allies.

Analysis

Analysis techniques depend on the types of data collected and on the questions asked by the process evaluation. The fundamental requirement of a process evaluation is to explain not only the who, what, and where of the program, but also why things happened as they did.

Quantitative techniques are generally applied to surveys and program records. These techniques include simple descriptive statistics, such as counts, tabulations, and measures of central tendency. Statistical tests, such as t-tests, chi-square tests, and multivariate analyses (e.g., conjoint analysis, regression analysis, factor and cluster analysis, and discrete-choice analysis), are also used.

Simple descriptive statistics are applicable to most survey and program data and even to some interview data. In fact, most who, what, and where questions are answered with counts and tabulations. Statistical tests and multivariate analyses, however, require careful construction of survey instruments and sample-selection procedures to assure that the appropriate data are available for analysis.

Qualitative techniques are used in all process evaluations to analyze interviews, focus groups, and program documents and to integrate quantitative data with these data sources. Qualitative analysis requires a structured and detailed organizational framework. Data are derived from a variety of sources requiring the analyst to systematically organize and classify the data into the analysis framework.

Miles and Huberman (1984) provide a detailed sourcebook on qualitative data-analysis methods; they describe three components of the analysis: data reduction, data display, and conclusion drawing and verification.

Data reduction transforms raw data into usable material. This reduction includes, for example, reviewing interviews to summarize and synthesize key issues and to formulate new questions. This process transforms the data into a presentation format from which conclusions can be drawn. For example, responses to a series of questions can be examined for key contact groups, or a matrix can be developed to tabulate responses across a variety of issues.

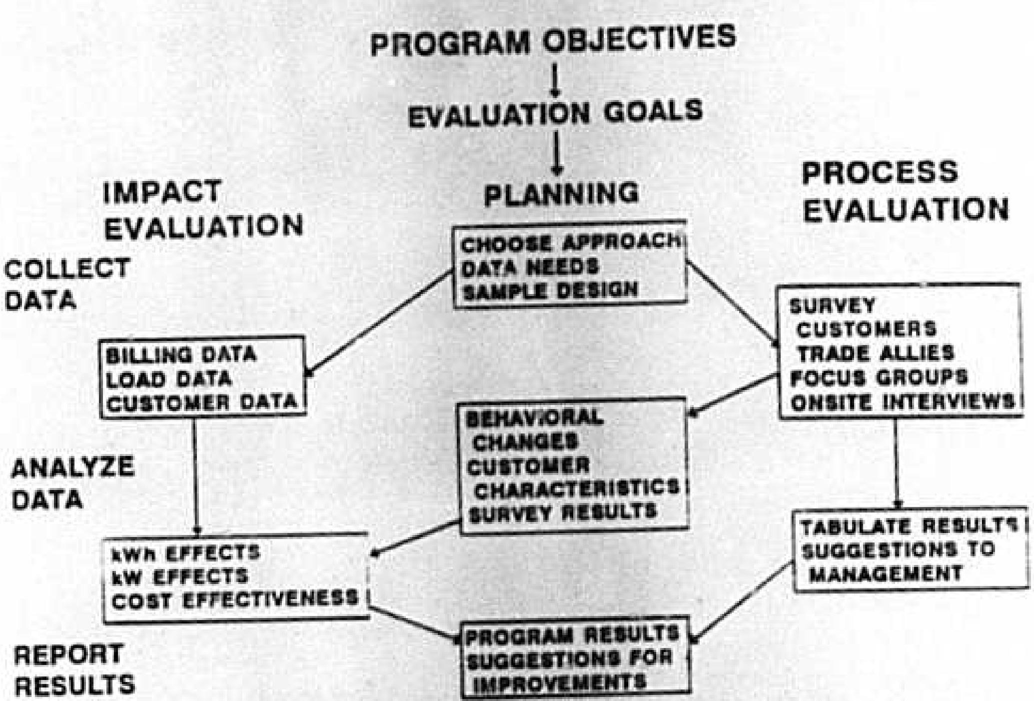


Fig. 18. **Although process and impact evaluations differ in their underlying purpose, they both deal with the behavioral changes that explain the observed outcomes and offer suggestions for program improvement.**

Because process evaluation personnel are "on the ground" early in the program, issues of data quality and availability can be addressed. Will the data needed for an impact evaluation be available in a timely manner? Evaluators assume that basic data (such as monthly electricity bills) are available for utility customers in recognizable form. This assumption is often incorrect. It is not unusual, for example, to find no data on customer contribution to the cost of DSM measures, to discover that entire classes of customer have meters read sporadically (with electricity use estimated the remainder of the time), to find billing histories archived in obscure formats, or to find that only a fraction of the data is digitized (and the rest exists only on paper at widely scattered district offices).

Often, no corporate memory of prior evaluations or data systems exists. Usable data are often discovered through the process evaluation. Early identification of problems and opportunities regarding data systems allows corrections to be made early enough to be useful in impact evaluation. Or it allows the impact evaluator enough time to change the evaluation strategy. For example, BPA's Commercial Audit Program evolved from a resource-acquisition program to a data-gathering/resource-assessment program to a field test of alternative auditing methods. Little documentation of these changes was made. The resulting database was limited in its usefulness when impact evaluations were implemented (Lerman and Bronfman 1986a; Cambridge Systematics and ERCE 1989).

Process and impact evaluations often address the same participant and nonparticipant customers. Impact evaluators may survey samples of participants and nonparticipants regarding decisions to participate, nonprogrammatic actions, building energy-use patterns, and equipment/appliance portfolios. Process evaluations may also deal with decisions to participate, participant and nonparticipant characteristics, attitudes, satisfaction with services, and intentions to take action. Many different purposes can be combined into one survey instrument. Again, with process evaluations in the field early, additional data can be gathered through follow-up surveys or site visits to make the data more current or to add information on new issues, such as measure life.

A few residential retrofit programs have been in operation for six to eight years. In situations like this, where year-to-year savings have been stable, impact evaluations may not be performed on an annual basis. Rather, periodic process evaluations are conducted, focusing on specific issues, such as market acceptance of changing incentives or measures offered. These "maintenance" evaluations and special projects can trigger new impact evaluations if results are substantially different from expectations.

Process evaluation has assumed a larger role in utility program evaluation, especially in settings where integrated resource planning and incentive regulation are in force. Utilities often do not have the time to pilot-test programs, to wait several years until the first impact results are available, or to conduct a priori market assessments or DSM-resource assessments. By looking at the goals of program evaluation, rather than the methods used to answer specific questions, we find that process evaluation is becoming a more universal tool in addressing utility DSM concerns.

Table 28 shows the key issues related to evaluations of DSM programs. The two main goals of evaluation are to estimate program savings and to optimize program design and delivery. Under each of these headings are the issues that utilities might examine. While several issues are addressed by one evaluation approach, others can be addressed by both process- and impact-evaluation methods. While energy-savings and net-savings analyses are addressed with only impact-evaluation methods, other issues (like persistence of savings and measure life) require both.

CONCLUSIONS

Process evaluations can help program planners, designers, and implementors in determining appropriate ways to improve DSM programs. Process evaluations also provide valuable information for deciding when a program is no longer required and when program efforts should be enhanced. Perhaps most importantly, such evaluations provide a vital tool for interpreting the results of impact evaluation.

Process evaluation, at its most basic, is the assessment of program implementation for the purpose of improving program delivery. With the increasing importance of DSM programs and the growing scrutiny by utility staff, regulators, and outside parties, process evaluations are becoming valuable tools because of their timing, flexibility, and easy integration into program operations.

Table 28. DSM-program-evaluation components and methods

	Impact evaluation	Process evaluation
Program savings		
Load-shape impacts	✓	
Total savings	✓	
Net savings	✓	
Attribution of effects	✓	
Persistence of savings	✓	✓
Cost-effectiveness	✓	✓
Program Optimization		
Infrastructure support		✓
Measure life	✓	✓
Program implementation		✓
Market assessment		✓
Technology assessment	✓	✓
Customer preference and behavior		✓

REFERENCES

Applied Energy Group, Inc., *Massachusetts Joint Utility End-Use Monitoring Project*, Final Report, Woodbury, NY, August, 1988.

Applied Energy Group, Inc. "Commercial Dollars and Sense Program Evaluation," *Impact Evaluation of Long Island Lighting Company's Electric Conservation and Load-Management Plan*, Woodbury, NY, 1989.

Applied Management Sciences, Inc. *Central Hudson Commercial/Industrial Incentives Program – Final Evaluation*, November, 1990a.

Applied Management Sciences, Inc. *Process Evaluation of Worcester Energy Fitness Program*, prepared for New England Electric System, Westborough, MA, March, 1990b.

Arkin, H. *Handbook of Sampling for Auditing and Accounting*, 2nd ed., McGraw-Hill, 1974.

Association of Demand-Side Management Professionals. "Are You Ready for DEEP Data?", *Strategies* 2(3), Berkeley, CA, Summer, 1991, p. 16.

Berry, L. *The Administrative Costs of Energy Conservation Programs*, ORNL/CON-294, Oak Ridge, TN: Oak Ridge National Laboratory, November, 1989.

Berry, L. *The Market Penetration of Energy-Efficiency Programs*, ORNL/CON-299, Oak Ridge, TN: Oak Ridge National Laboratory, April, 1990.

Beschen, D. A. and M. A. Brown. *Evaluation Plan for the Weatherization Assistance Program*, ORNL/TM-11668/V1, Oak Ridge, TN: Oak Ridge National Laboratory, August, 1991.

Birnbaum, I. and T. Davis. "Using Commercials to Sell DSM Programs: A Measurement of Customer Response," *Proceedings, Energy Program Evaluation: Conservation and Resource Management*, Argonne, Il: Argonne National Laboratory, 1989, pp. 195-200.

Blevins, R. P. and B. A. Miller. *1987 Survey of Commercial-Sector Demand-Side Management Programs*, EPRI CU-6294, Palo Alto, CA: Electric Power Research Institute, March, 1989a.

Blevins, R. P. and B. A. Miller. *1988 Survey of Residential-Sector Demand-Side Management Programs*, EPRI CU-6546, Palo Alto, CA: Electric Power Research Institute, October, 1989b.

Bonneville Power Administration (BPA). *Guidelines for Proposing an Energy Savings Plan Efficiency Project*, Portland, OR: Bonneville Power Administration, July, 1988.

Bonneville Power Administration (BPA). *Evaluation of the Commercial Incentives Pilot Program Report*. Portland. OR: Cambridge Systematics, 1990.

Boston Edison. *1991 Demand-Side Management Monitoring and Evaluation Plan*, Boston. MA. July, 1991.

Bradford, P. Remarks at Workshop on Demand-Side Management Program Evaluation. sponsored by New York State Energy Research and Development Authority and New York State Electric & Gas Corp., Saratoga Springs, NY. March. 1991.

Brian Gard William Lesh Inc. *Report on Market Research and Program Recommendations. Bonneville Power Administration Regionwide Promotion of Energy Efficient Appliances*. January, 1986.

Brooker. R. and R. Fichtner. "Measuring Program Impact in a Commercial and Industrial Lighting Program: The Case of Wisconsin Electric's Smart Money Program." *Proceedings of the 1991 International Energy Evaluation Conference*. Argonne. IL: Argonne National Laboratory, August. 1991. pp. 94-99.

Brown. M.A., J. O. Kolb, D. Baylon, M. H. Haeri, and D. L. White. *The Impact of Bonneville's Model Conservation Standards on the Energy Efficiency of New Home Construction*, ORNL/CON-310, Oak Ridge, TN: Oak Ridge National Laboratory, August, 1991.

Brugger, D. and R. Brooker. "Evaluation of Wisconsin Electric's Dealer Incentive Plan," *Proceedings of the 1991 International Energy Evaluation Conference*, Argonne, IL: Argonne National Laboratory, August, 1991, pp. 290-293.

Calder. B. "Cognitive Consistency and Consumer Behavior," *Perspectives in Consumer Behavior*. Harold H. Kassarjian and Thomas S. Robertson (eds.), Glenview, Il: Scott. Forseman and Company, 1973.

California Public Utilities Commission and California Energy Commission. *Standard Practice Manual for Cost-Benefit Analysis of Demand-Side Management Programs*, San Francisco, CA and Sacramento, CA, December. 1987.

Cambridge Systematic and ERCE. *Commercial Incentive Pilot Program (CIPP) Process Evaluation*, Berkeley, CA. 1989.

Campbell. D. T. and L. C. Stanley. *Experimental and Quasi-Experimental Designs for Research*. Skokie. IL: Rand McNally, 1966.

Central Hudson Gas and Electric Company. *Comprehensive Process and Impact Evaluation of the Dollar Savers Program for Commercial Lighting, Air Conditioning, Motors, and Anything Goes Subprograms*. Boulder. CO. December. 1991.

Energy Program Evaluation Conference. "Energy Program Evaluation: Uses, Methods and Results." *Proceedings 1991 International Energy Program Evaluation Conference*. CONF-910807. Argonne. IL: Argonne National Laboratory, August, 1991.

Environmental and Energy Services Company and Pacific Energy Associates. *Evaluation Plan for the Orange and Rockland Utilities Demand Side Bidding Programs*. Portland, OR, 1989 (Revised 1990).

Evans. P. and Peters. J. "Integration of Process Evaluation and Program Design," *Proceedings, Energy Program Evaluation: Conservation and Resource Management*, Argonne, IL: Argonne National Laboratory, 1989, pp. 115-119.

Faruqui, A. et al. *Impact of Demand-Side Management on Future Customer Electricity Demand: An Update*. EPRI CU-6953, Palo Alto, CA: Electric Power Research Institute, September, 1990.

Feldman. S. "Forging the Link: Needs Segmentation and Geodemographics." *Proceedings, 1988 EPRI/EUMRC Market Research Symposium*, Scottsdale, AZ, November, 1988.

Fels. M. "The Princeton Scorekeeping Method: An Introduction, *Energy and Buildings*, 9, Feb./May 1986, pp. 5-18.

Fels, M. and C. Reynolds. "Toward Standardizing the Measurement of Whole-Building Energy Savings in DSM Programs," *Proceedings of the 1991 International Energy Program Evaluation Conference*, Chicago, IL, August 1991.

Franklin, J. and J. Thrasher. *An Introduction to Program Evaluation*, New York: John Wiley and Sons, 1976.

Freeman Research Resources. *A Process Evaluation of ENERGY INITIATIVE*. Vol. I. Monterey, MA, May, 1991.

Gallagher, J. T. *DSM Incentives in New York State: A Critique of Initial Utility Methods*, Albany, NY: New York Department of Public Service, March, 1991.

Gordon. F. M., M. McRae, M. Rufo, and D. Baylon. "Use of Commercial Energy Efficiency Measure Service Life Estimates in Program and Resource Planning," *Proceedings, ACEEE 1988 Summer Study on Energy Efficiency in Buildings*, Washington, DC: American Council for an Energy Efficient Economy, Vol. 3, August, 1988, pp. 3.84 – 3.97.

Haeri, M. H. *Electricity Savings Three Years after Participation in the BPA Regionwide Weatherization Program*, prepared for the Bonneville Power Administration. Portland, OR, September, 1988.

Hanser, P. Electric Power Research Institute, personal communication, 1991.

Heckman, J. J. "Dummy Endogenous Variables in a Simultaneous Equation System." *Econometrica* 46(6), 1978, pp. 931-959.

Heckman, J. J. and R. Robb. "Alternative Methods for Evaluating the Impact of Interventions," *Longitudinal Analysis of Labor Market Data*, J. J. Heckman and B. Singer, (eds), New York: Cambridge University Press, 1985, pp. 156-246.

Hickman, C. and T. Steele. "Site Visits to Enhance Commercial Program Evaluation." *Proceedings of 1991 International Energy Program Evaluation Conference*, Argonne, Il: Argonne National Laboratory, August, 1991, pp. 174-180.

Hicks, E. "Conservation for Profit and Program Evaluation in a Collaborative Planning Process." *Proceedings from the ACEEE 1990 Summer Study on Energy Efficiency in Buildings*, Washington, DC: American Council for an Energy Efficient Economy, 1990, pp. 6.47.

Hirst, E. "Evaluating Demand-Side Management Programs." *Electric Perspectives 14*(6), Nov./Dec., 1990a, pp. 24-30.

Hirst, E. *Balancing the Scales: Data on Electricity Supply and Demand*, ORNL/CON-306, Oak Ridge, TN: Oak Ridge National Laboratory, September, 1990b.

Hirst, E. *Possible Effects of Utility DSM Programs, 1990 to 2010*, ORNL/CON-312, Oak Ridge, TN: Oak Ridge National Laboratory, January, 1991.

Hirst, E. and K. M. Keating. "Dynamics of Energy Savings Due to Conservation Programs." *Energy Systems and Policy 10*(3), 1987, pp. 257-273.

Hirst, E. and C. Sabo. *Electric-Utility DSM-Program Terminology*, ORNL/CON-337, Oak Ridge, TN: Oak Ridge National Laboratory, October, 1991.

Hirst, E., D. L. White, and R. Goeltz. *Three Years after Participation: Electricity Savings Due to the BPA Residential Weatherization Pilot Program*, ORNL/CON-166, Oak Ridge, TN: Oak Ridge National Laboratory, January, 1985.

Horowitz, M. J., L. K. Ecker, and P. Degens. *Long Term Impacts of the Interim Residential Weatherization Program on Household Energy Savings*, (ERCE/DSM-65), prepared for the Bonneville Power Administration, Portland, OR, May, 1991.

Joint Committee on Standards for Educational Evaluation. *Standards for Evaluation of Educational Programs, Projects, and Methods*, New York: McGraw-Hill, 1981.

Kaplan, M., B. Jones, and J. Jansen. "DOE2.1C Model Calibration with Monitored End-Use Data." *Proceedings of the ACEEE 1990 Summer Study on Energy Efficiency in Buildings*, Washington, DC: American Council for an Energy Efficient Economy, August, 1990, pp. 10.115 – 10.125.

Keating, K. M. "Self-Selection Bias: Are We Beating a Dead Horse?" *Evaluation and Program Planning*, 12(2), 1989, pp. 137-142.

Keating, K. "Issue Paper on Free Riders." *1990 Conservation and Load-Management Evaluation Plan*, New England Electric System, June, 1990.

Keating, K. Personal communication, Bonneville Power Administration, Portland, OR, September, 1991.

Keating, K. and S. Blachman. "In Search of an Impact: An Evaluation of an Institutional Buildings Program." *Proceedings, Energy Conservation Program Evaluation, Practical Methods, Useful Results, 1987 Conference*. Vol. 1, Argonne, IL: Argonne National Laboratory, August 1987, pp. 107-116.

Keating, K. and E. Hicks. (eds). "Program Evaluation." *Proceedings, ACEEE 1990 Summer Study on Energy Efficiency in Buildings*, Vol. 6, Washington, DC: American Council for an Energy Efficient Economy, August, 1990.

Keating, K. M. and E. Hirst. "Advantages and Limits of Longitudinal Evaluation Research in Energy Conservation," *Proceedings, Evaluation and Program Planning*, Vol. 9, Washington, DC: American Council for an Energy Efficient Economy, August, 1986, pp. 113-120.

Keating, K. M. and T. V. Oliver. *Evaluating Commercial Conservation Programs: A Difficult Business*, Bonneville Power Administration, Portland, OR (draft), 1991.

King, M. and W. Gavelis. "Informed Conditional Demand Analysis: An Integrating Methodology for Estimating Energy Impacts," *Proceedings, American Council for an Energy Efficient Economy, Washington, DC*, August, 1990, pp. 10.127 - 10.134.

Komor, P., W. Kempton, and J. Haberl. *Energy Use, Information, and Behavior in Small Commercial Buildings*, Princeton, NJ: Center for Energy and Environmental Studies, Princeton University, July, 1989.

Krause, F. and J. Eto. *Least-Cost Utility Planning Handbook for Public Utility Commissioners, The Demand Side: Conceptual and Methodological Issues*, Vol. 2, Washington, DC: National Association of Regulatory Utility Commissioners, December, 1988.

Kreitler, V. "Market Issues in Free Rider Estimation," *Proceedings of the Electric Utility Marketing Research Council EUMRC/EPRI Conference*, EPRI CU-7010, Bala Cynwyd, PA: Synergic Resources Corporation, November, 1990, pp. 10-28, 33, 34.

Kreitler, V. "On Customer Choice and Free Ridership in Utility Programs," *Proceedings of the 1991 International Energy Evaluation Conference*, Argonne, IL: Argonne National Laboratory, August, 1991, pp. 303.

Kunkle, R. *Case Study Analysis of the Institutional Buildings Program in Washington State*, Olympia, WA: Washington State Energy Office, October, 1990.

Kunkle, R. and C. Johnson. "Building Changes and Conservation Measure Problems: Institutional Buildings Program Case Studies," prepared for the Bonneville Power Administration, Portland, OR, May, 1991.

Kushler, M. G. "Use of Evaluation to Improve Energy Conservation Programs: A Review and Case Study," *Journal of Social Issues* 45(1), 1989, pp. 153-168.

Lerman, D. *Regional Study of Residential Water Heating Equipment*, IEAL/PO-24, Portland, OR: Bonneville Power Administration, 1987.

Lerman, D. and B. Bronfman. *Process Evaluation of the Bonneville Power Administration Long Term Residential Conservation Program*. ERC/PO-5, Portland, OR: Bonneville Power Administration, 1984.

Lerman, D. and B. Bronfman. *Process Evaluation of the BPA Commercial Audit Program*. IEAL/PO-4, Portland, OR: Bonneville Power Administration, 1986a.

Lerman, D. and B. Bronfman. *Process Evaluation of the Tacoma MCS Adoption: Part II*, IEAL/PO-9, Portland, OR: Bonneville Power Administration, 1986b.

Lerman, D., B. Bronfman, and B. Tonn. *Process Evaluation of the BPA Residential Weatherization Pilot Program*, ORNL/CON-138, Oak Ridge, TN: Oak Ridge National Laboratory, 1983.

Long Island Lighting Company. *Electric Class of Customer Study*, Hicksville, NY, 1972.

Long Island Lighting Company and Applied Energy Group, Inc. *Impact Evaluation of Long Island Lighting Company's 1990 Electric Conservation and Load-Management Plan*, Hicksville, NY, 1991.

Lui, D. and J. Fang. *Free Rider Ratios in Conservation programs: Estimates and Issues*, staff topic paper prepared for the Illinois Commerce Commission, Springfield, IL, 1989.

Lui, D. and J. Fang. "Issues in Free Rider Estimations," *Proceedings of the 1990 ACEEE Summer Study on Energy Efficiency in Buildings*, Vol. 6, Washington, DC: American Council for an Energy Efficient Economy, August, 1990, pp. 6.96.

MacDonald, S. L. *Residential Peak Load-Reduction Project, Shed LMG Test Analysis for 1986 Operating Season*, San Francisco, CA: Pacific Gas & Electric Company, July, 1987.

Malcolm, W. H. "Ten Years After: Lessons Learned from the Residential Direct Control Program." *Proceedings: Demand-Side Management Strategies for the 90s, Fourth National Conference on Utility DSM Programs.* EPRI CU-6367, Palo Alto, CA: Electric Power Research Institute, April, 1989, pp. 30-1-8.

Massachusetts Electric. *1990 DSM Performance Measurement Report*, Westborough, MA: New England Electric System, June, 1991.

McRae, M. "Don't Shoot the Messenger: Conducting DSM Process Evaluations in (Inevitably) Political Environments," *Proceedings, ACEEE 1990 Summer Study on Energy Efficiency in Buildings, Program Evaluation*, Vol. 6, Washington, DC: American Council for an Energy Efficient Economy, August, 1990, pp. 6.99-6.107.

Meier, A. *Supply Curves of Conserved Energy*, Ph.D. dissertation, Berkeley, CA: Energy and Resources Group, University of California, May, 1982.

Miles, M. and A. Huberman. *Qualitative Data Analysis: A Sourcebook of New Methods.* Newbury Park, CA: Sage Publications, 1984.

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

Nadel, S. "Use of Simple Performance Indices to Help Guide Review of DSM Program Performance," *Proceedings, National Conference on Integrated Resource Planning*, Washington, DC: National Association of Regulatory Utility Commissioners, April, 1991, pp. 116-130.

Nadel, S., K. Cress, and M. Ticknor. "The NEES Stand-By Generation Program: Design, Results, and Lessons Learned," *Demand-Side Management Strategies for the 90s, Proceedings, Fourth National Conference on Utility DSM Programs*, CU-6367, Vol. 2, Palo Alto, CA: Electric Power Research Institute, April, 1989, pp. 65-1–14.

Nadel, S. M. and K. M. Keating. "Engineering Estimates vs. Impact Evaluation Results: How Do They Compare and Why?" *Proceedings of 1991 International Energy Program Evaluation Conference, CONF-910807*, Argonne, Il: Argonne National Laboratory, August, 1991, pp. 24-33.

Nadel, S. and H. Tress. *The Achievable Conservation Potential in New York State from Utility Demand-Side Management Programs*, Report 90-18, Albany, NY: New York State Energy Research and Development Authority, November, 1990.

New England Electric System (NEES). *Six Month Evaluation of New England Electric System Partners in Energy Planning Programs*, Westborough, MA: New England Electric System, April, 1988.

New England Power Service Company, *1991 DSM Performance Measurement Report*, Westboro, Ma, June, 1991.

New Jersey Conservation Assessment Team. *Final Contractor's Report to the New Jersey Conservation Analysis Team,*" Boulder, CO: RCG/Hagler, Bailly, Inc., August, 1990.

New York Power Pool and the Empire State Electric Energy Research Corporation, *Report of Member Electric Systems pursuant to Article VIII, Section 149-b of the Public Service Law*, 1976, Vol. 1, pp. 96-171.

New York Public Service Commission. *Order Concerning 1991 and 1992 Demand Side Management Plans*, Case 28223, Albany, NY, November, 1990.

New York State Energy Office. *Program Evaluation of the New York State Energy Efficient Appliance Rebate Demonstration Program*, Albany, NY, May, 1989.

Nichols. D., et al. *Savings from the Smart Money Program, An Audit of the 125 Megawatt Demand Reduction*, Boston, MA: The Tellus Institute, May, 1990.

Northeast Region Demand-Side Management Data Exchange. *Program Data Summary Report*, Washington, DC: Edison Electric Institute, June, 1989.

Northeast Utilities. *ESLR Short-Duration Monitoring Interim Report*, Cambridge, MA: RLW Analytics, Inc., 1991a.

Northeast Utilities. *Integrating Information for DSM Evaluation Using the Engineering Calibration Approach*, Cambridge, MA: RLW Analytics, Inc., 1991b.

Okumo, D. "Multifamily Retrofit Electricity Savings: the Seattle City Light Experience," *Proceedings of the ACEEE 1990 Summer Study on Efficiency in Buildings*, Washington, DC: American Council for an Energy Efficient Economy, 1990, p. 6.119.

Ozog, M. *Free Ridership in WP&L Buysmart Residential DSM Programs*, Draft report prepared for Wisconsin Power and Light, July, 1991.

Pacific Consulting Group, Inc. *Evaluation of PG&E's 1984-85 Customized Rebate Program Final Report*, Vol. I, June, 1986.

Parti, C. and M. Parti. "The Total and Appliance-Specific Conditional Demand for Electricity in the Household Sector," *Bell Journal of Economics*, Spring, 1980, pp. 309-321.

Patton, M. Q. *Practical Evaluation*, Beverly Hills, CA: Sage Publications, 1982a.

Patton, M. Q. *Creative Evaluation*, Newbury Park, CA: Sage Publications, 1982b.

Patton, M. Q. *How to Use Qualitative Methods in Evaluation*. Newbury Park, CA: Sage Publications, 1987, (Fifth printing 1990).

Peters, J., D. Barry, M. Horowitz, and F. Gordon. *Evaluation of the Orange and Rockland Utilities Competitive Bidding Program for Demand Side Resources*. Interim Report. ERCE/DSM-66. Portland, OR: Bonneville Power Administration, 1990.

Peters, J. and B. Bronfman. *Street and Area Lighting Program Evaluation Update*, IEAL/PO-6, Portland, OR: Bonneville Power Administration, 1986.

Peters, J. and G. Gustafson. *Process Evaluation of the Sponsor-Designed Site Specific Program*, Portland, OR: Bonneville Power Administration, March, 1987.

Peters, J., G. Gustafson and M. Vowles. *Process Evaluation of the BPA Purchase of Energy Savings Pilot Program*, IEAL/PO-20, Portland, OR: Bonneville Power Administration, 1987.

Peters, J., H. Haeri, and G. Gustafson. *Comparative Analysis of Commercial Sector Financing Mechanisms: CIPP, PES Pilot and Puget Incentive Programs*. ERC/PO-35. Portland, OR: Bonneville Power Administration, 1988.

Peters, J., H. Haeri, and P. Seratt. *Evaluation of the PG&E Home Energy Survey Programs*, ERCE/PO-54, San Francisco, CA: Pacific Gas & Electric Corp., 1990.

Peters, J., K. Oswald, and M. Horowitz. *Evaluation of the Northeast Utilities EnergyCHECK Program*, Hartford, CT: Northeast Utilities, December, 1990.

Petersen, F. J. "Remodel and Tenancy Changes: Threats to the Reliability of Commercial Conservation Savings," *Proceedings, 1990 ACEEE Summer Study on Energy Efficiency in Buildings*, Vol. 3. Washington, DC: American Council for an Energy Efficient Economy, August, 1990, pp. 3.165 – 3.172.

Petersen, F. J. and S. Sandler. "Institutional Buildings in Flux," *Proceedings of 1991 International Energy Program Evaluation Conference*, Argonne, Il: Argonne National Laboratory, August, 1991, pp. 41-48.

Pollard, W. E. "Bayesian Statistics for Evaluation Research: An Introduction," *Contemporary Evaluation Research Series No. 8*, Newbury Park, CA: Sage Publications, 1986.

Prahl, R. "Evaluation for PUC's," *Proceedings of the 1988 Summer Study on Energy Efficiency in Buildings*, Washington, DC: American Council for an Energy Efficient Economy, 1988, p. 9.126.

Puget Energy Services, Inc. Final Report for the *Commercial Retrofit End-Use Study*, prepared for the Bonneville Power Administration, Portland, OR, March, 1991.

Regional Economic Research,Inc. *SDG&E Non-Residential Audit Evaluation*, October, 1990.

Regional Economic Research. Inc. *An Evaluation of Niagara Mohawk's Low Cost Measures Program*, June, 1991.

Rochester Gas & Electric and Applied Energy Group, Inc. *Evaluation of Rochester Gas & Electric's 1990 Demand Side Management Programs*, 1991.

Rogers, E. "Evaluation of a Residential Appliance Rebate Program Using Billing Record Analyses." *Proceedings, 1989 Energy Program Evaluation Conference*. Argonne. IL: Argonne National Laboratory, 1989.

Rossi. P. and F. Freeman. *Evaluation – A Systematic Approach*. Beverly Hills, CA: Sage Publications, 1982.

San Diego Gas & Electric Company. *SDG&E Non-Residential Audit Evaluation*. San Diego, CA: Regional Economic Research. Inc., 1991.

Saxonis. W. "Program Evaluation of the New York State Energy Efficient Appliance Rebate Demonstration Program: The Free Rider Factor," *Proceedings of the 1989 International Energy Program Evaluation Conference: Conservation and Resource Management*, Argonne, IL: Argonne National Laboratory, August, 1989, pp. 237-240.

Saxonis, W. "Measuring Free Riders: Does the Economic Climate Make a Difference?" *Proceedings of the 1991 International Energy Evaluation Conference*, Argonne, IL: Argonne National Laboratory, August, 1991, pp. 286.

Schmitt, S.A. *Measuring Uncertainty: An Elementary Introduction to Bayesian Statistics*, Reading, MA: Addison-Wesley, 1969.

Schoch-McDaniel. K. *Third Year Analysis of Electricity Use and Savings for the Hood River Conservation Project*, prepared for the Bonneville Power Administration. Portland, OR, October, 1990.

Schön, A. and K. Hamilton. "Commercial Sector Alternatives to End-Use Metering," *Proceedings of End-Use Load Information and Its Role in DSM Conference*, Irvine, CA, 1990.

Skumatz, L. A., K. M. Lorberau, R. J. Moe, R. J. Bordner, and R. D. Chandler. *Bonneville Measure Life Study: Effect of Commercial Building Changes on Energy Using Equipment*, prepared for the Bonneville Power Administration, Portland, OR, 1991.

Stout, T. and W. Gilmore. "Motor Incentive Programs: Promoting Premium Efficiency Motors," *Demand-Side Management: Partnerships in Planning, Proceedings ECNE National Conference on Utility DSM Programs*, EPRI CU-6598, Bedford, MA: Electric Council of New England. November, 1989, pp. 27-1-11.

Stucky, L., M. Khawaja, S. French, C. Collette, and H. G. Peach. *Impact Evaluation of the Low-Income Segment of the Weather Shield and Attic Attack Program*, Augusta, ME: Central Maine Power, 1990.

Sumi, D. H. and B. Coates. *Longitudinal Evaluation of Energy Savings from Seattle City Light's Home Energy Loan Program*, Seattle WA: Seattle City Light, December, 1988.

Synergic Resources Corporation. *Analysis of the NYSEG Power Pincher High Efficiency Refrigerator Program*, Bala Cynwyd, PA: Synergic Resources Corporation, November, 1987.

Synergic Resources Corporation. *Process Evaluation of the Bonneville Long Term Weatherization Program: 1987-1989*, SRC Report 7582-R3, Bala Cynwyd, PA: Synergic Resources Corporation. 1991.

TechPlan Associates. *Evaluation of the Easy Savers Program*, Poughkeepsie, NY: Central Hudson Gas and Electric Corporation. March, 1990.

Tolkin, E. *Energy Fitness Resurvey Analysis*, memo report to New England Electric System, Westborough, MA, October, 1990.

Townsley, M. and R. Wright. "Measuring DSM Impacts: End-Use Metering and the Engineering Calibration Approach," *Proceedings of 2nd Conference on End-Use Load Information and Its Role in DSM Association of DSM Professionals*, Ervine, CA, 1990.

Train, K., P. Ignelzi, and Kumm. "Evaluation of a Conservation Program for Commercial and Industrial Customers," *Energy*, 10(10), 1985, pp. 1079-88.

Train, K. E. *Correcting Self-Selection Bias in the Estimation of Audit Program Impacts*. Berkeley, CA: Cambridge Systematics, 1988.

Train, K. and P. Ignelzi. "The Economic Value of Energy-Saving Investments by Commercial and Industrial Firms," *Energy, 12*(7), 1987.

Van Liere, K. and T. McKinnell. "Comparison of Alternative Segmentation Approaches for Residential Customers," *Proceedings Energy Program Evaluation: Conservation and Resource Management*, Argonne, IL: Argonne National Laboratory, 1989, pp. 457-464.

Vine, E., O. De Buen, and C. Goldman. *Stimulating Utilities to Promote Energy Efficiency: Process Evaluation of Madison Gas and Electric's Competition Pilot Program*, Report LBL-29642, Berkeley, CA: Lawrence Berkeley Laboratory, December, 1990.

Violette, D. "Use of Metered Data in DSM Evaluation." *Proceedings of End-Use Load Information and Its Role in DSM*, sponsored by Aloha Systems, Inc., Irvine, CA: The Fleming Group and the Association of DSM Professionals, June 1990.

Violette, D. and M. Ozog. "Correction for Self-Selection Bias: Theory and Application," *Energy Program Evaluation: Conservation and Resource Management Conference*, Argonne, IL: Argonne National Laboratory, August. 1989, pp. 241-250.

Violette, D., M. Ozog, and G. Wear. "Measuring Free Ridership: Do Some Experimental Designs Control Twice for Free Ridership?" *Proceedings of the 1991 International Energy Evaluation Conference*, Argonne, IL: Argonne National Laboratory, August. 1991, pp. 296.

Violette, D., M. Ozog, M. Keneipp, and F. Stern. *Impact Evaluation of Demand-Side Management Programs, Vol. 1: A Guide to Current Practice*, CU-7179, Palo Alto, CA: Electric Power Research Institute, February, 1991.

Walpole, R. E. and R. H. Myers. *Probability and Statistics for Engineers and Scientists*, 3rd ed, McMi"·n Publications, 1985.

Weiss, C. H. *Evaluation Research: Methods of Assessing Program Effectiveness*. Englewood Cliffs, NJ: Prentice Hall, 1972. P. 11.

White, D. L. and M. A. Brown. *Electricity Savings Among Participants Three Years after Weatherization in Bonneville's 1986 Residential Weatherization Program*, ORNL/CON-305, Oak Ridge, TN: Oak Ridge National Laboratory, September, 1990.

White, D. L., T. K. Stovall, and B. E. Tonn. *Persistence of the Impact of the Hood River Conservation Project on Typical and Peak Loads: Three Years after Weatherization*, ORNL/CON-321, Oak Ridge, TN: Oak Ridge National Laboratory, June, 1991.

Wiggins, D. San Diego Gas and Electric, personal communication, February, 1991.

Wisconsin Electric. *1990 Annual Evaluation Report for Wisconsin Electric's Demand-Side Management Programs*. March. 1991.

Wisconsin Public Service Corporation. *Good Cents Home Program: Impacts and Evaluation*, Boulder, CO: RCG/Hagler, Bailly, Inc., March, 1989.

XENERGY, Inc. *Demand-side Management Program Evaluation Scoping Study*, ESEERCO Project EP90-34, New York, NY: Empire State Energy Research Corporation, November, 1990.

XENERGY, Inc. *Joint Utility Baseline New Construction Practices Study*, Burlington, MA. 1991.

ANNOTATED BIBLIOGRAPHY

John Reed

INTRODUCTION

This bibliography directs readers to works that are seminal to evaluation and energy evaluation. Many of the items are infrequently cited in the energy literature but provide important foundations on which the field of energy evaluation is built. No attempt has been made to make an exhaustive list. For example, many of the high-quality studies from the ACEEE Summer Studies series, the Chicago Evaluation Conference Proceedings, and EPRI conference proceedings are underrepresented. We urge the reader to explore those sources for examples of high-quality evaluations. The brief annotations are designed to help readers locate those books and articles that will meet their needs. In addition to being chosen for their importance, these works were also chosen for accessibility. Most of these items should be available in university libraries or through interlibrary loan. The list is presented alphabetically according to the name of the first author under three headings: General Evaluation Methods, Analytical Methods, and Energy Evaluation.

GENERAL EVALUATION METHODS

Cook, T. D., and D. T. Campbell. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*, Chicago, IL: Rand McNally, 1979.

Energy-program evaluators are seldom able to conduct evaluations with classical experimental design (i.e., measurements taken before and after program implementation for a control group and an experimental group with random assignment of subjects to the two groups). Most frequently, energy researchers are faced with situations in which they are asked to perform an evaluation after a program has been implemented and for which no control group is available. This volume, which is the successor to an earlier volume by Stanley and Campbell, *Experimental and Quasi-Experimental Designs*, discusses just such problems. The book begins with an excellent discussion of validity and then moves on to discussions of quasi-experimental designs, such as the one-group post-program design, the post-program only with nonequivalent groups design, and the one-group pretest-posttest design. Other designs are also discussed. The book portrays a variety of statistical techniques that may be used to reduce uncertainty about causality given the use of quasi-experimental designs. Beginners in evaluation and even the mature researcher can benefit from reviewing this book from time to time.

Dillman, D. A. *Mail and Telephone Surveys, The Total Design Method*, New York, NY: John Wiley and Sons, 1978.

The development, administration, and analysis of surveys is often poorly done because evaluators underestimate the difficulties and complexities of doing a quality job. Want to know how to word a question? Want to reduce the complexity of a question? Want to know how to improve the response rate for a mailed survey? Want to know how to reduce the cost of processing data once a questionnaire is returned? This volume provides answers to these and other questions that may confront an evaluator. Highly recommended even for the seasoned veteran of survey research, who can benefit from perusing this volume to see if all the bases have been covered before that questionnaire is mailed or the telephone survey begun.

Patton, M. Q. *Creative Evaluation*, Newbury Park, CA: Sage Publications, 1982.

This is not your typical cookbook providing a recipe for how to do evaluation. Rather, Patton says that the book contains "ideas, techniques, and approaches" that he "never taught ... in formal courses" and that still are probably not a part of most courses in evaluation. What Patton does is provide methods that he uses to creatively work with decision makers and users of evaluation information. The book describes activities like flow charting, picture thinking, story thinking, and simulation games. If you have trouble communicating about evaluation with management and other decision makers, this volume will help. Evaluation reports might be less dry and better utilized if some of these techniques were used.

Patton, M. Q. *How to Use Qualitative Methods in Evaluation*, Newbury Park, CA: Sage Publications, 1987.

In-depth interviews, direct observation, and the analysis of written documents, such as speeches, diaries, program records, and open-ended written items on questionnaires, are the main types of qualitative methods. Decision makers often view the type of information produced by qualitative methods with suspicion. However, if qualitative methods are applied with knowledge and skill, the result can be reliable and valid data that provide insight not available from quantitative methods. As the author puts it, "Systematic and rigorous observation involves far more than just being present and looking around. Skillful interviewing involves much more than just asking questions. Content analysis requires considerably more than just reading to see what's there." This book describes how and when to use qualitative methods and how to conduct observation and in-depth interviews and how to analyze written documents as well as the data produced by observation and interviews. The techniques described here are particularly useful for scoping a problem and for process evaluation. They also may help in interpreting the anomalies that occur in quantitative data.

Patton, M. Q. *Practical Evaluation*, Beverly Hills, CA: Sage Publications, 1982.

This is another volume that is full of practical guidance. Patton's motivation for writing this book was to call attention to the need for excellence in evaluation. The book contains the "thoughtful" chapters, "Thoughtful Questionnaires," "Thoughtful Interviewing," and "Thoughtful Methods Decisions." It also contains a chapter on "Managing Management Information Systems," which discusses the problem of managing data, which is often overlooked, is little discussed, and can lead to severe problems. While management-information science has changed since this volume was written, this chapter is a helpful reminder, and the information is still of value.

The book is very readable. Instead of being prescriptive, the author provides examples that lead the reader to draw his own conclusions. For example, at one point, the author gives examples of three styles of tabular presentation. Reading the text is almost unnecessary to grasp the point the author is making about presenting tabular data. This book contains important reading, but unlike many books on evaluation, this is one that can be read with enjoyment and understanding.

Shadish, W. R., Jr., T. D. Cook, and L. C. Leviton. *Foundations of Program Evaluation: Theories of Practice*, Newbury Park, CA: Sage Publications, 1991.

This new book discusses the work of some of the most respected names in evaluation, Campbell, Weiss, Wholey, Cronbach, and Rossi to name a few. The book has several purposes. First, it provides an intellectual history of evaluation so that the reader can understand how and why evaluation practices have evolved. Second, the book attempts to broaden the perspective of evaluation practitioners by highlighting how the historical circumstances in various sectors (education and mental health, for example), have shaped the form that evaluations in those sectors have taken. This is useful because it helps to enlarge our view of what evaluation is and the options that are available to practitioners. It also helps to sensitize the practitioner to the ways in which the context of a problem may shape the evaluation. Third, the book is targeted to policy-makers, who often make decisions about evaluation without understanding evaluation and without consulting evaluation specialists.

Energy evaluation as a specialty has been somewhat insular from the main stream of evaluation as a discipline. At times, it seems as if energy-evaluation practitioners have been quick to reinvent the wheel and slow to recognize and to solve problems that have already been dealt with in other sectors. The exposure of energy evaluators to this book will partially remedy this problem.

Weiss, C. *Evaluation Research: Methods of Assessing Program Effectiveness*. Englewood Cliffs, NJ: Prentice Hall, 1972.

The focus of this slim volume is the application of evaluation research methods to real-world environments. The author offers guidance on adapting "textbook methods" to the reality of the world in which evaluation must be done. One of the chapters, "Formulating the Question and Measuring the Answer," focuses on the problem of developing the goals of a program and how these relate to measuring program outcomes. Another chapter discusses the changing nature of programs and the relations between the evaluator(s) and program personnel and the impact of these on evaluation. Also, a chapter treats the use of evaluation results. In contrast to the Cook and Campbell volume with its theoretical focus, this volume deals with the practical issues of evaluation. Those new to the field might find this a good introduction and useful first reading before tackling Cook and Campbell.

Wholey, J. S. *Evaluation and Effective Public Management*. Boston, MA: Little Brown and Company, 1983.

The premise that underlies this book is that more managerial and analytic attention needs to be given to program assessment and improvement. Although this book is oriented to government programs and the civil service, the content is of value to those conducting energy evaluations for utilities as well as public service commissions. The book is divided into four parts. Two of these parts are of particular interest. One is the section on getting agreement on results-oriented objectives. Many programs and many evaluations fail because of lack of agreement among participants about objectives. This section includes a chapter discussing the problem of setting realistic and measurable objectives. The section on "Assessing Program Performance and Results" discusses the problems of developing performance measures and evaluation designs. Also in this section is a discussion of outcome monitoring. The book contains numerous examples of results-oriented management. This book is of value to managers wanting to know how to effectively use evaluations to better manage their programs.

STATISTICAL AND ANALYTICAL METHODS

Aigner, D. J., C. Sorooshian, and P. Kerwin. "Conditional Demand Analysis for Estimating Residential End-Use Load Profiles," *The Energy Journal*, 5(3), 1984, pp. 81-97.

One of the real frustrations for energy evaluators is obtaining locale-specific end-use load shapes. Although costs are dropping, end-use load data are still costly to gather and analyze. This article describes the use of conditional-demand analysis (see also the article by Parti and Parti) to statistically extract end-use load profiles from whole-household data. Twenty-four regression equations were fitted with the hourly consumption of electricity averaged over the days of the month regressed on the presence or absence of household appliances. The result was a series of well-defined residential load shapes, although the "load level often seemed questionable." The authors attribute this to the simple model specification and to the fact that ownership and intensity of use were not incorporated into the model. Nonetheless, the technique is well illustrated. With ownership and intensity of use incorporated into the model, the results should be excellent. Conditional-demand analysis offers a lower-cost alternative to end-use metering, especially if a utility already has remote-meter-reading capability in place.

Fels, M. F. (ed). "Measuring Energy Savings: The Scorekeeping Approach," a special edition of *Energy and Buildings*, 9(1), 1986.

One cannot be in the business of energy evaluation for long without having to compare energy-consumption measurements taken at different times (before and after a program) or places (that are influenced by different weather conditions). Examples of such situations include comparing energy consumption for the same household in two different years and comparing the consumption of different households with different monthly billing cycles. Fortunately, adjusting the data to remove the effects of weather is straightforward. The required data include whole-building meter readings and daily average temperature readings. This volume, which is an excellent treatment of weather normalization, presents a series of papers that use and discuss PRISM. There is an excellent introduction to the subject by the editor. Many of the remaining 15 papers present examples of the use of PRISM for different building types (single-family and multifamily dwellings) and fuel types (electricity, wood, and oil). Additional papers comment on the reliability of PRISM and other uses of this model. If you evaluate programs with aggregate building data, this is an important collection of papers.

Loether, H. J., and D. G. McTavish. *Descriptive and Inferential Statistics*, Boston, MA: Allyn and Bacon, Inc., 1976.

This is one of the best of the general surveys of statistics and one of the easiest to use. The book does a good job of presenting the techniques so that they are easily understood. Although it lacks discussion of some of the most recent advances, it covers most of the descriptive and inferential statistics that an evaluator might use. It has a particularly strong section on the analysis of cross-classification data, which is useful if you are analyzing questionnaire or household data. It also has a section on graphic presentation and analysis that may be helpful if you use a microcomputer graphics software package to analyze or display data. The book also does a good job of helping the user to understand when to use the various statistical techniques. The book contains numerous sidebars that summarize important material. The book is highly recommended if you are new to the use of statistics or if you need a quick refresher.

Mosteller, F., and J. W. Tukey. *Data Analysis and Regression: A Second Course in Statistics*, Reading, MA: Addison-Wesley, 1977.

The key to this book is its subtitle, "A Second Course in Statistics." This is a good follow-on to the Tukey book on Exploratory Data Analysis and the Loether and McTavish volumes described elsewhere in this bibliography. This is a book about regression analysis and not a general statistics book. The book discusses what regression is and what regression coefficients can and cannot tell us. It discusses model fitting and what information can be extracted from residuals. Most of us were introduced to statistics based on the normal distribution. However, the energy researcher might find the beta distribution more useful, especially for modelling the distribution of duty cycle of appliances, such as air conditioners. The first chapter in this book discusses the strengths and weaknesses of the Gaussian (normal) and beta distributions. If you need to use regression or to evaluate studies that used regression, this text should be very useful.

Parti, M., and C. Parti. "The Total and Appliance-Specific Conditional Demand for Electricity in the Household Sector," *Bell Journal of Economics*, 2(2), Spring, 1980, pp. 309-321.

Faced with the need to determine appliance-specific energy consumption, the evaluator can use engineering estimates, install end-use monitoring devices, or use conditional-demand analysis. With appropriate models, engineering estimates are straightforward although often not as accurate as one would like, and the estimates tend not reflect to regional differences in appliances and usage patterns. End-use monitoring provides excellent data but is expensive. An alternative is conditional-demand analysis, especially if your utility already has the data required for the analysis. Using information about household-specific appliance stocks, one can disaggregate total household load into estimates of energy consumption for each appliance. The term "conditional" comes from the notion that you either have an appliance or do not. Regression techniques, in which the availability of appliances are represented as dummy variables, are used to estimate the parameters of the component demand functions. Add demographic variables, information about the size of the building and the season, follow the recipe in this article, and you, too, can estimate energy usage by appliance. For a slightly different use of this approach, see the Aigner article above.

Tukey, J. W. *Exploratory Data Analysis*, Reading, MA: Addison-Wesley, 1977.

This is the classic work on Exploratory Data Analysis (EDA). EDA is a series of techniques that allows one to look effectively at data. The purpose of EDA is to gain insight into data when the patterns in the data or the meaning of data may be unclear. This volume describes techniques that allow the user to look at data in new ways. Most of the techniques (for example, leaf-and-stem analysis) can be applied with pencil and paper, although statistical packages for microcomputers will now present the same data.

Violette, D., M. Ozog, M. Keneipp, and F. Stern. *Impact Evaluation of Demand-Side Management Programs: A Guide to Current Practice*, Vol. 1, CU-7179, Palo Alto, CA: EPRI, 1991.

As its title implies, this volume focuses on impact evaluation. While it discusses many general evaluation issues, it is dedicated to presenting the state of the art in impact analysis. Important chapters on data collection, engineering methods, statistical methods, and combining data from multiple sources are included. Each chapter discusses specific techniques in detail as well as providing examples. The strengths and weaknesses of the various techniques are also discussed. If you need detailed information on specific impact-analysis techniques, this volume is the place to start. (The reader should be forewarned that this an EPRI document and that its use is governed by a license agreement. This may limit its usefulness to organizations that are affiliated with EPRI.) Most of the material in this volume is available in other formats, but access is not nearly so convenient.

ENERGY EVALUATIONS

ACEEE 1990 Summer Study on Energy Efficiency in Buildings, Washington, DC: American Council for an Energy Efficient Economy, 1990.

These are the proceedings for the ACEEE 1990 Summer Study. The general foci of these proceedings are buildings and energy technologies, but many other issues (such as organizational behavior and decision making, evaluations of incentive programs, social stratification and appliance stratification, and efficiency and greenhouse gases) are discussed. In particular, one volume of the proceedings deals exclusively with the evaluation of DSM programs. This conference alternates years with the Chicago Evaluation Conference. If you want to know the state of the art in energy-efficient technologies, this is the place to start. Proceedings for 1988 and 1986 are also available.

Hirst, E. "The Hood River Conservation Project: An Evaluators Dream," *Evaluation Review*, 12(3), June, 1988, pp. 310-325.

The Hood River Conservation Project was a 5-year, $20-million research and demonstration project to install cost-effective conservation measures in electrically heated homes in Hood River, Oregon. The measures were aimed at the building shell to reduce electricity use for SH and at WH retrofits. This article summarizes the remarkable and largely successful efforts to design a comprehensive evaluation to address energy-policy issues important to the Pacific Northwest. Evaluation results are presented concerning the number of eligible households that participated, the number of recommended conservation actions that were adopted, and the actual electricity use and savings that were achieved.

Energy Program Evaluation: Uses, Methods, and Results: Proceedings of the 1991 International Energy Program Evaluation Conference, CONF-910807, Chicago, IL: National Energy Program Evaluation Conference, August, 1991.

The International Energy Program Evaluation Conference is a biannual conference that alternates with the ACEEE Summer Study. The papers in this proceedings include evaluations of energy programs, discussions of evaluation methods, and discussions of the practical problems of evaluating energy programs. This proceedings and its predecessor volumes present the state of the art in energy program evaluation.

Nadel, S. M., and K. M. Keating. "Engineering Estimates vs. Impact Evaluation Results: How Do They Compare and Why?" in *Energy Program Evaluation: Uses, Methods, and Results: Proceedings of the 1991 International Energy Program Evaluation Conference*, CONF-910807. Chicago, IL: National Energy Program Evaluation Conference, August, 1991, pp. 24-33.

How good are engineering estimates compared to the results of impact evaluations based on billing or metered data? This paper examines the results of 42 studies in an attempt to answer this question. Impact-evaluation results are lower than engineering estimates for residential retrofit programs, commercial and residential lighting programs, and low-flow showerhead programs; and they are about equivalent for residential appliance programs, new construction programs, and multiple-measure C&I programs. The reasons for the discrepancies are erroneous assumptions in the engineering calculations, complex interactions among measures, quality-control problems with the measures, and greater-than-expected adoption of measures by nonparticipants. This work is an example of a metastudy that can help guide better program designs and evaluations.

Nadel, S. M., and M. Ticknor. "Electricity Savings from a Small Commercial & Industrial Lighting Retrofit Program: Approaches and Results," in *Energy Program Evaluation: Conservation and Resource Management: Proceedings of the 1989 International Evaluation Conference*, Chicago, IL: National Energy Program Evaluation Conference, 1989, pp. 107-111.

This short piece describes an impact evaluation of a pilot lighting retrofit program for small C&I customers offered by the Massachusetts Electric Company during 1985 to 1987. This study is unusual because four methods were employed: engineering calculations, comparison of pre- and postprogram daily kWh use with a control group, comparison of pre- and postprogram daily kWh use with a survey on changes in participants' energy-use patterns, and conditional-demand analysis. The first and third methods were of questionable accuracy, while the second worked well, and the fourth performed adequately but not without some difficulties. This paper is a good comparison of the problems and promises of various methods.

Sumi, D. H., and B. Coates. "Persistence of Energy Savings in Seattle City Light's Residential Weatherization Program." *Proceedings of the 1989 International Energy Program Evaluation Conference: Conservation and Resource Management*, Chicago, IL: National Energy Program Evaluation Conference, 1989, pp. 311-316.

Because energy evaluation and DSM programs are relatively young, few studies deal with long-term impacts of energy programs. This study is one of the few examples. This paper is about the persistence of net impacts and not the persistence of savings from an installed measure. The durability of net savings over the study years 1982-1987 was about 73%. The study is well done, and the authors note some important limitations to their work, including the lack of data on household characteristics that might have changed the estimates of net impacts. If you are interested in the long-term impact of savings, you should study this paper.

Train, K., P. Ignelzi, and M. Kumm. "Evaluation of a Conservation Program for Commercial and Industrial Customers." *Energy*, 10(10), pp. 1079-88.

This work studies the savings resulting from a conservation program directed at C&I customers in the Southern California Edison service area. After an energy audit, rebates were offered to customers for installing items like time clocks, photocells, load controllers, lighting-system changes, and skylights. The evaluation found that the cost per kWh of savings was less than the average of fuel costs of generation, indicating that the measures were cost-effective. This study is important because it is one of the early studies of C&I conservation and because of the techniques used to conduct the evaluation. Program participation was estimated with logit models. Savings estimates were based on regression techniques that took into account nonweather factors influencing load (hours of operation, square footage, etc.), weather-sensitive loads, and conservation-induced effects.

AUTHOR BIOGRAPHIES

Benson H. Bronfman is President of Bronfman & Associates, a consulting practice specializing in DSM program planning and evaluation. He is an expert in evaluation planning, process-evaluation theory and methods, and the application of research design and statistical approaches to analyzing DSM program impacts. His experience includes evaluations of DSM programs in all sectors and at all stages of development. He previously headed the DSM evaluation practice at ERC International and the Social Impact Analysis Group at Oak Ridge National Laboratory. He holds a Ph.D. in Political Science from the University of Oregon.

George L. Fitzpatrick is the President of Applied Energy Group, Inc., a technical consulting and software-development firm serving the electric-utility industry, primarily in DSM evaluation and implementation, load research, strategic planning, power-plant cost and performance analysis, utility-performance-standard development, and comparative economic analysis of utility investments. Prior to founding Applied Energy Group, Inc. in 1981, he was Vice President at Stone & Webster Management Consultants, Inc. He was also Manager of the Load- Research, Costing, and Forecast Division at Long Island Lighting Company. He holds an M.B.A. and a B.A. in Economics, both from St. John's University.

Elizabeth G. Hicks is the Director of Demand Planning at New England Power Service Company. She oversees the planning, design, and evaluation of all DSM programs, and the research and development on new DSM technologies. Previously, she was the Manager of Conservation and Load-Management Planning and Evaluation, responsible for the development of DSM programs, participation in collaborative planning processes, and evaluation of DSM programs. She has a B.A. in Economics and Mathematics from the College of William and Mary and an M.S. in Resource Systems and Policy Design from Dartmouth College.

Eric Hirst is a Corporate Fellow at Oak Ridge National Laboratory. He received a Ph.D. in Mechanical Engineering from Stanford University in 1968. Since 1970, Hirst has been at ORNL. He directs ORNL's work on integrated resource planning, which includes projects on uncertainty, modeling, collaboratives, and evaluation of DSM programs.

Marc G. Hoffman, Vice President at XENERGY, is a resource economist specializing in utility planning and evaluation with 11 years of experience in supply- and demand-side planning. For several years, he has developed leading applications of end-use analysis to utilities' load-forecasting and demand-side-planning needs. Most recently, he is leading the development of XENERGY's evaluation efforts. He holds an M.S. in Agricultural Economics from the University of Connecticut and a B.A. from Brandeis University.

AUTHOR BIOGRAPHIES

Benson H. Bronfman is President of Bronfman & Associates, a consulting practice specializing in DSM program planning and evaluation. He is an expert in evaluation planning, process-evaluation theory and methods, and the application of research design and statistical approaches to analyzing DSM program impacts. His experience includes evaluations of DSM programs in all sectors and at all stages of development. He previously headed the DSM evaluation practice at ERC International and the Social Impact Analysis Group at Oak Ridge National Laboratory. He holds a Ph.D. in Political Science from the University of Oregon.

George L. Fitzpatrick is the President of Applied Energy Group, Inc., a technical consulting and software-development firm serving the electric-utility industry, primarily in DSM evaluation and implementation, load research, strategic planning, power-plant cost and performance analysis, utility-performance-standard development, and comparative economic analysis of utility investments. Prior to founding Applied Energy Group, Inc. in 1981, he was Vice President at Stone & Webster Management Consultants, Inc. He was also Manager of the Load- Research, Costing, and Forecast Division at Long Island Lighting Company. He holds an M.B.A. and a B.A. in Economics, both from St. John's University.

Elizabeth G. Hicks is the Director of Demand Planning at New England Power Service Company. She oversees the planning, design, and evaluation of all DSM programs, and the research and development on new DSM technologies. Previously, she was the Manager of Conservation and Load-Management Planning and Evaluation, responsible for the development of DSM programs, participation in collaborative planning processes, and evaluation of DSM programs. She has a B.A. in Economics and Mathematics from the College of William and Mary and an M.S. in Resource Systems and Policy Design from Dartmouth College.

Eric Hirst is a Corporate Fellow at Oak Ridge National Laboratory. He received a Ph.D. in Mechanical Engineering from Stanford University in 1968. Since 1970, Hirst has been at ORNL. He directs ORNL's work on integrated resource planning, which includes projects on uncertainty, modeling, collaboratives, and evaluation of DSM programs.

Marc G. Hoffman, Vice President at XENERGY, is a resource economist specializing in utility planning and evaluation with 11 years of experience in supply- and demand-side planning. For several years, he has developed leading applications of end-use analysis to utilities' load-forecasting and demand-side-planning needs. Most recently, he is leading the development of XENERGY's evaluation efforts. He holds an M.S. in Agricultural Economics from the University of Connecticut and a B.A. from Brandeis University.

Kenneth Keating began his career in energy program evaluation at the Bonneville Power Administration two weeks after receiving his Ph.D. from Washington State University in 1982. Previously, he had worked on evaluations of traffic programs and solar energy. During his 10 years at Bonneville, he has been responsible for more than 140 evaluation projects. He has been active on the planning committees for both the ACEEE Summer Study and the five International Conferences on Energy Program Evaluation, which, in 1991, honored him for his work in the development of the field.

Harvey Michaels, Senior Vice President at XENERGY, is responsible for the company's R&D in end-use methods and software as well as for corporate operational management. In his 13 years at XENERGY, he has led a number of significant projects in end-use forecasting and conservation analysis, including the development of industrial energy-management guidebooks for 15 trade-association energy programs, development of an energy-auditing system, and an end-use research system. He has an M.S. in Economics and Policy and a B.S. in Civil Engineering from Massachusetts Institute of Technology.

Steven Nadel is a Senior Associate with the ACEEE. Nadel's areas of emphasis are programs, policies, and technologies to reduce electricity demand. This work focuses on utility DSM efforts and on the efficiency of appliances, lighting, and motor systems. Prior to joining ACEEE, he spent two years planning and evaluating-electricity conservation programs for the New England Electric System. He has a B.S. in Government, an M.A. in Environmental Studies from Wesleyan University, and an M.S. in Energy Management from New York Institute of Technology.

Jane S. Peters is a Project Director with Barakat & Chamberlin, Inc., and co-manager of the Portland, Oregon, office. She specializes in both impact and process evaluation of DSM programs. She is an expert in process and market evaluations focusing on customer response and on the effectiveness of program delivery and administrative mechanisms. She received her Ph.D. in Urban Studies from Portland State University and her A.B. in Psychology from Occidental College.

John Reed is group leader of the Energy and Environmental Applications Group, Energy Division, Oak Ridge National Laboratory. He has done considerable work in the evaluation of energy-conservation programs. He helped design and conduct a series of workshops on program evaluation for state energy offices. He worked on the load-management portion of the Athens Automation and Control experiment in Athens, Tennessee. Most recently, he has been involved in the design of computer-based decision-support systems. He received a Ph.D. in sociology from Cornell University.

William Saxonis is Manager of the New York State Energy Office's Evaluation Unit, which is responsible for evaluating a wide range of energy-conservation programs. He is the author of several articles dealing with free-rider measurement and evaluation planning. He currently serves on the planning committee of the International Energy Program Evaluation Conference and leads a project team to evaluate the long-term effectiveness of energy retrofits in multifamily buildings. He earned an M.A. in Political Science from the Rockefeller College of Public Affairs and Policy.

Andrew Schön, Vice President of DSM Engineering at XENERGY since 1979, is the architect of a hybrid engineering and statistical methodology to develop end-use load shapes. He has served as project director for numerous end-use research, evaluation, and engineering studies of DSM technologies. Additionally, he oversees and directs all engineering design and monitoring services. Mr. Schön has a B.S. in Mechanical Engineering from the University of Massachusetts and an M.S. in Technology and Policy from the Massachusetts Institute of Technology.

Daniel M. Violette is Senior Vice President of RCG/Hagler, Bailly, Inc. He has extensive experience in the evaluation, design, testing, and implementation of DSM programs. He was the lead author of *Impact Evaluation of DSM Programs* and has managed a number of large evaluations of DSM program impacts. He has developed and used econometric, engineering, and hybrid models to best use limited and uncertain information. He is a founding member of the Association of Demand-Side Management Professionals, and he currently chairs the Association's committee on DSM program monitoring and evaluation. He has a Ph.D. in Economics from the University of Colorado.

INTERNAL DISTRIBUTION

1. V. D. Baxter
2. L. Berry
3. D. S. Bjornstad
4. R. B. Braid
5. M. A. Brown
6. R. S. Carlsmith
7. R. Cantor
8. F. C. Chen
9. J. Christian
10. G. Courville
11. S. J. Dale
12. P. D. Fairchild
13. W. Fulkerson
14. M. B. Gettings
15. C. W. Hagan
16. L. J. Hill
17. E. Hillsman
18. E. Hirst
19. P. J. Hughes
20. C. R. Kerley
21. J. O. Kolb
22. M. A. Kuliasha
23. R. Lee
24. P. Leiby
25. J. M. MacDonald
26. L. N. McCold
27. V. C. Mei
28. W. R. Mixon
29. S. Purucker
30. S. Rayner
31. J. H. Reed
32. D. E. Reichle
33. D. T. Rizy
34. M. Schweitzer
35. R. B. Shelton
36. G. G. Stevenson
37. J. N. Stone
38. J. Van Dyke
39. J. M. Veigel (ORAU)
40. D. B. Waddle
41. D. L. White
42. T. J. Wilbanks
43. ORNL Patent Office
44. Central Research Library
45. Document Reference Section
46. Laboratory Records (RC)

47-49. Laboratory Records Dept.

EXTERNAL DISTRIBUTION

50. Bruce G. Buchanan, Computer Science Department, University of Pittsburgh, 206 Mineral Industries Building, Pittsburgh, PA 15260

51. Allan Hirsch, Vice President, Environmental Sciences, and Director, Washington Operations, 5109 Leesburg Pike, Suite 414, Falls Church, VA 22041

52. Martin Williams, Professor, Department of Economics, Northern Illinois University, DeKalb, IL 60115

53. Helen Ingram, Director, Udall Center for Studies in Public Policy, University of Arizona, 803/811 East First Street, Tucson, AZ 85719

54. Calvin MacCracken, President, Calmac Manufacturing Corporation, 101 West Sheffield Avenue, P. O. Box 710, Englewood, NJ 07631

54.-64. OSTI, U. S. Department of Energy, P. O. Box 62, Oak Ridge, Tennessee 37831

65. Office of Assistant Manager for Energy Research and Development, DOE/ORO, P. O. Box 2001 Oak Ridge, TN 37831-8600

66.-1050. External Energy Efficiency and Renewables Section Distribution Mailing List and extra copies to E.M. Schorn, 4500N, H-19A

INTERNAL DISTRIBUTION

1. V. D. Baxter
2. L. Berry
3. D. S. Bjornstad
4. R. B. Braid
5. M. A. Brown
6. R. S. Carlsmith
7. R. Cantor
8. F. C. Chen
9. J. Christian
10. G. Courville
11. S. J. Dale
12. P. D. Fairchild
13. W. Fulkerson
14. M. B. Gettings
15. C. W. Hagan
16. L. J. Hill
17. E. Hillsman
18. E. Hirst
19. P. J. Hughes
20. C. R. Kerley
21. J. O. Kolb
22. M. A. Kuliasha
23. R. Lee
24. P. Leiby
25. J. M. MacDonald
26. L. N. McCold
27. V. C. Mei
28. W. R. Mixon
29. S. Purucker
30. S. Rayner
31. J. H. Reed
32. D. E. Reichle
33. D. T. Rizy
34. M. Schweitzer
35. R. B. Shelton
36. G. G. Stevenson
37. J. N. Stone
38. J. Van Dyke
39. J. M. Veigel (ORAU)
40. D. B. Waddle
41. D. L. White
42. T. J. Wilbanks
43. ORNL Patent Office
44. Central Research Library
45. Document Reference Section
46. Laboratory Records (RC)

47-49. Laboratory Records Dept.

EXTERNAL DISTRIBUTION

50. Bruce G. Buchanan, Computer Science Department, University of Pittsburgh, 206 Mineral Industries Building, Pittsburgh, PA 15260

51. Allan Hirsch, Vice President, Environmental Sciences, and Director, Washington Operations, 5109 Leesburg Pike, Suite 414, Falls Church, VA 22041

52. Martin Williams, Professor, Department of Economics, Northern Illinois University, DeKalb, IL 60115

53. Helen Ingram, Director, Udall Center for Studies in Public Policy, University of Arizona, 803/811 East First Street, Tucson, AZ 85719

54. Calvin MacCracken, President, Calmac Manufacturing Corporation, 101 West Sheffield Avenue, P. O. Box 710, Englewood, NJ 07631

54.-64. OSTI, U. S. Department of Energy, P. O. Box 62, Oak Ridge, Tennessee 37831

65. Office of Assistant Manager for Energy Research and Development, DOE/ORO, P. O. Box 2001 Oak Ridge, TN 37831-8600

66.-1050. External Energy Efficiency and Renewables Section Distribution Mailing List and extra copies to E.M. Schorn, 4500N, H-19A

MONITORING AND EVALUATION PLAN

Prepared for

IOWA-ILLINOIS GAS AND ELECTRIC COMPANY
Davenport, Iowa

Prepared by

XENERGY Inc.
Madison, Wisconsin
and
Burlington, Massachusetts

July 1992

XENERGY Inc. • Three Burlington Woods • Burlington, MA 01803-4543 • (617) 273-5'00 • FAX (617) 229-4867

XENERGY

MONITORING AND EVALUATION PLAN

Prepared for

IOWA-ILLINOIS GAS AND ELECTRIC COMPANY
Davenport, Iowa

Prepared by

XENERGY Inc.
Madison, Wisconsin
and
Burlington, Massachusetts

July 1992

TABLE OF CONTENTS

This document describes Iowa-Illinois' Monitoring and Evaluation Plan for its energy efficiency programs.

1.1 Purpose of the Evaluation

Iowa-Illinois' monitoring and evaluation efforts have two primary objectives. The first is to review program markets and implementation, from start-up through maturity, identifying effective strategies for program delivery and discovering barriers which require attention. The studies that support these tasks are known as market assessments and process evaluations. The second major objective is to measure program benefits related to energy (kilowatt-hour and therm) savings and capacity (kilowatt and peak day therm) reductions; studies of this type are referred to as impact evaluations.

It is important to recognize the distinction between the terms *monitoring* and *evaluation*. Typically, monitoring refers to on-going tracking of the performance of a program. Utility staff oversee the activity of the program such as the number of participants, the type of equipment installed under the program, amount of rebates, length of time to process a rebate check, etc. Program monitoring occurs on a routine basis, i.e., daily, weekly or monthly. Evaluation refers to assessing the performance of the program over a longer but discrete period of time, such as a calendar year. The program evaluation consists of market, process and impact analyses. Designing program plans which include both monitoring and evaluation activities ensures a comprehensive assessment of the program on a short-term and long-term basis.

Formal Evaluation Goals. This Monitoring and Evaluation Plan is a follow-up to Iowa-Illinois' energy efficiency plan which was approved by the Iowa Utilities Board in an order issued January 29, 1992. Iowa-Illinois has contracted with XENERGY Inc. of Burlington, Massachusetts, and Madison, Wisconsin, to develop this evaluation plan independent of any other program design and development initiatives. In this manner Iowa-Illinois is assured of an objective assessment of its program compliance with the rule making specified by IAC 199-35.1 through 35.10. The development of specific monitoring and evaluation plans is guided by the following directives:

- Monitor and evaluate programs for the duration of the implementation period plus two years. [IAC 199-35.8(11)]

- Monitor program performance. Create procedures for a data base to track program implementation including monthly customer contacts, installations, and expenditures. Compare program progress to program goals and schedules. Determine customer conservation installations and practices. Assess operation arrangements and program management. [IAC 199-35.8(11)"a"(1),(2) and (3)]

- Generate a full range of utility and customer cost data to determine actual cost effectiveness of programs and program components. Document actual installed measure and O&M costs, including customer and utility costs. Document utility overhead, marketing, incentives, administration, monitoring and evaluation, and staff costs. Document savings, measure costs and estimated O&M costs projected for specific projects. Determine cost effectiveness of technologies and technology packages. [IAC 199-35.8(11)"a"(4)]

- Provide quick feedback for interim program assessment where necessary and to support ongoing improvements to program design and implementation. Develop short-term evaluation plans, where appropriate, that provide indicators of program direction and magnitude of savings to insure that programs are performing as anticipated. Develop schedules for intermediate and long-term data gathering and reporting activities. [IAC 199-35.8(11)"a"(5)]

- Assess consumer decision-making to refine program delivery mechanisms and reach target segments in a cost-effective manner. Coordinate program evaluation with market research to determine the components of customer decision-making for energy efficiency investments. Identify opportunities and barriers for specific market segments to optimize program delivery systems and develop reasonable expectations for program market penetration. [IAC 199-35.8(11)"b"]

- Determine the demand reductions and energy savings attributable to the energy efficiency programs. Collect, process and analyze data to determine the level of demand reduction and energy saving acquired through specific programs, and where possible by market segment, local area, building type and energy efficiency measure. [IAC 199-35.8(11)"b"(1)]

- Determine hourly load shapes. Collect, process and analyze data to determine the impact of conservation measures and programs on customer, end-use, market segment, local area and system load shapes. [IAC 199-38.5(11)"b"(1)]

- Determine the reliability of demand reductions and energy savings over time. Gather data over time after treatment to determine the persistence of reductions and savings. Consider external influences such as weather and economic activity. Consider direct influences of occupant behavior, structural and equipment operation changes which affect electricity use and energy efficiency measure performance. [IAC 199-35.8(11)"b"(2),(3)]

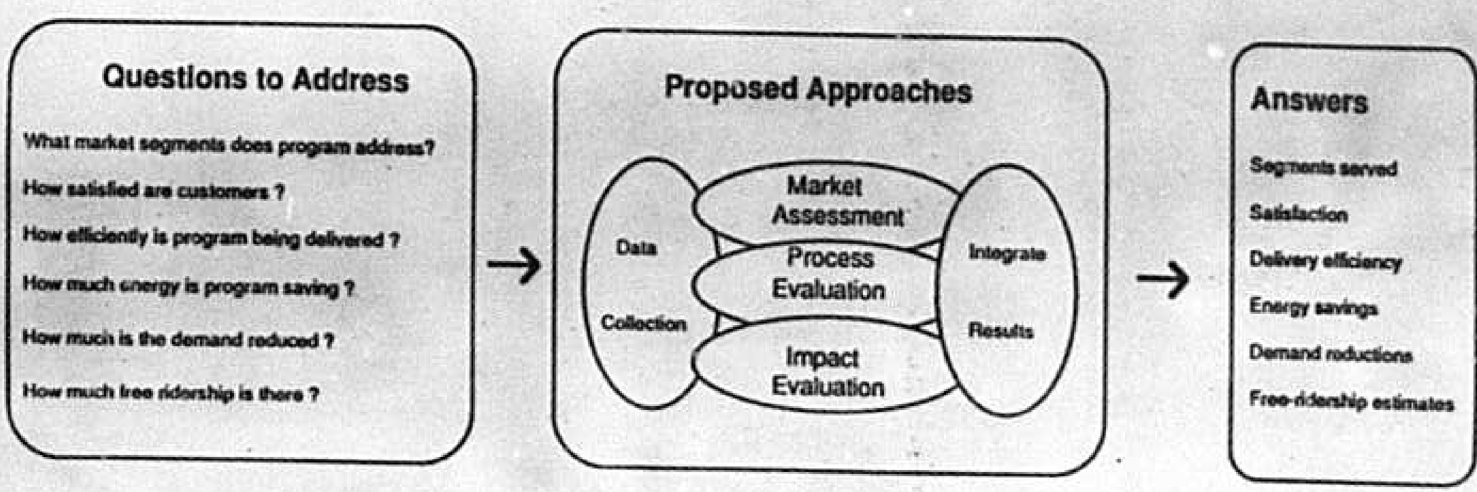
- Develop monitoring and evaluation schedules that are coincident with program implementation activities. [IAC 199-35.8(11)"b"(6)]

1.2 General Approach

The monitoring and evaluation plans presented are comprehensive and recognize the complementary nature of market assessments, process evaluations and impact evaluations. The approaches described in this plan have been chosen to answer specific questions for each program. As shown in Figure 1-1, these approaches function in a coordinated manner. Some common data collection activities provide data that may be used by one or more of the three assessment and evaluation activities. Then the results of the market assessment, process evaluation and impact evaluation are integrated and presented.

One of the primary areas of interplay between process and impact evaluation is data collection via surveys. Survey data are necessary to support impact analysis because they provide detailed information on customers' microdata that allows the investigators to control for the differential effects of customer characteristics when statistically estimating energy savings. Surveys also provide key

Figure 1-1
Overview of Program Evaluation



data for assessing issues such as program satisfaction and free drivers. These issues are important in determining the effectiveness of program administration and delivery for process evaluations and for adjusting gross impacts. Coordination of process and impact analyses requires identifying data needs on a program-by-program basis, and developing data collection strategies that meet these needs.

The evaluation plans proposed for Iowa-Illinois exhibit three key characteristics:

- Flexibility to facilitate adapting to program changes and additional evaluation requirements.
- Multiple impact evaluation approaches that complement engineering estimates and billing analysis with sample metering and building simulation where appropriate.
- Outcomes designed to support optimizing Iowa-Illinois' energy efficiency efforts and filing for program cost recovery.

To assure that the impacts of Iowa-Illinois' programs are monitored and evaluated carefully, two significant preliminary efforts are included in these plans. Both efforts involve engineering estimates of the impacts of the 1992 programs. An initial "shake down" analysis will be performed in September 1992 to ensure that tracking data are being properly collected and stored. Then a full "evaluation implementation analysis" for the 1992 programs will be conducted by March 1, 1993. These engineering estimates will produce gross savings figures for programs and individual measures.

The evaluation plans for each program include, in addition to these preliminary implementation analyses, complete long-term plans for impact evaluations.

1.3 Timeframe

The program evaluation time-frame for this plan falls into three periods. The implementation phase will occur in early 1992. Late 1992 and 1993 is a ramp-up period to the maximum participation levels. 1994 and beyond will represent program years where persistence and degradation of energy savings may become evident.

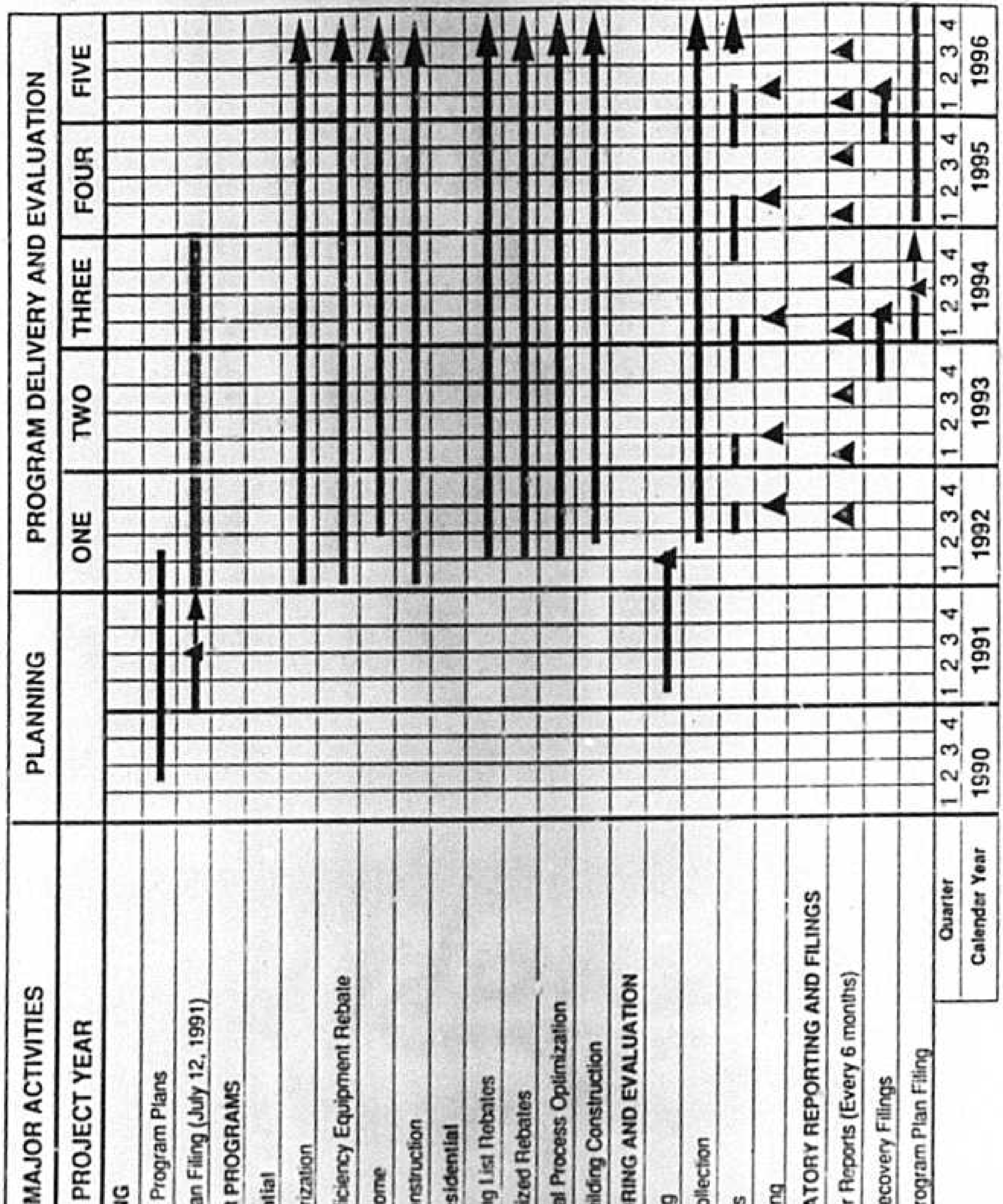
The overall timing of program delivery and evaluation is shown in Figure 1-2.

For 1992, the plan focuses on developing market, process and impact evaluation methodologies for all programs. The first critical part of this effort will be exploring the usefulness of available data resources, such as the customer billing files. The second critical effort will be reviewing Iowa-Illinois' tracking systems for both residential and non-residential programs. Engineering algorithms for estimating the impacts of individual measures will be developed during this time, further detailing the structure and depth of the tracking data base. Process assessment methodologies will begin within the third quarter of 1992 to encompass the initial start-up of program implementation at Iowa-Illinois.

During 1992 and 1993, baseline data will be collected and preliminary implementation analyses will provide impact estimates for all programs. The baseline data will be captured in the intermediate period in which some programs are fully implemented and others have just been launched. It will be important during this period for data collection efforts to be specific enough to meet the immediate requirements of programs in the field yet broad enough to accommodate those being implemented. Billing data for participant and non-participant samples will be examined to assure that procedures to identify appropriate records and transfer data are working. Metering will be installed and checked to assure data collection and processing steps are all functioning. Refinements to the database and evaluation plans will be made as experience is gained in implementing energy-efficiency programs. During the intermediate period, an initial impact assessment will be conducted to provide Iowa-Illinois with early information on the effectiveness of its energy efficiency programs. This preliminary assessment will involve early efforts in September 1992 to validate procedures and a full-scale implementation analysis to provide results by March 1, 1993.

In 1994 and beyond, the plan relies to a greater degree on executing methods which will be driven by levels of participation. These methods, although presented conceptually at this point, provide the blueprint for rigorous scientific studies that offer empirical estimates of program impacts. These will include statistical billing analyses, metering and simulation studies.

Figure 1-2
Timing of Program Delivery and Major Evaluation Activities



To support Iowa-Illinois in their near-term and long-term planning efforts and regulatory filings, program evaluation reports presenting the results of the analyses will be prepared. Table 1-1 describes the results that will be included in each report.

Periodic Review of Evaluation Plans. Specific items contained in the evaluation plans will be subject to frequent periodic review once the evaluation projects begin. These items include sample sizes, the use of site visits, evaluation schedules, and general approaches and methodologies for the analysis. There are several reasons for this further review and development, many of which hinge on information not currently available, or the likelihood of changing situations.

Changing situations could include wide variations in actual versus projected program participation rates, shifting and emerging implementation or evaluation objectives and issues, or field situations and evaluation costs which may be different from the expected. For example, if program participation is significantly higher than expected, then the planned sample sizes for customer surveys may no longer be appropriate. In this case, sample sizes would be revised and resulting analyses based on the latest participation rates. If the cost of site visits is significantly higher than anticipated, then Iowa-Illinois could reconsider the usefulness and the scope of the site visits. If field situations or general market conditions exist which are unexpected and significantly affect or alter program implementation, then Iowa-Illinois and XENERGY can revise the affected evaluation plan to address these changing conditions.

1.4 Outline of this Report

Following this introduction (Section 1), the questions to be addressed by the evaluation and the approaches the plan will take to answering them, are presented in Section 2. Proposed budgets for the different components of the evaluation are given in Section 3.

A separate section is then devoted to each of the individual programs to be evaluated (Sections 4 through 11). These sections provide the program-specific plan details.

Table 1-1
Summary of Evaluation Reports

	Oct. 92	Mar. 93	Mar. 94	Mar. 95	Mar. 96
Filing with IUB			*		*
Impact Results					
Adequacy of Tracking Database Data	X	X	√	√	
Engineering Estimates of Program Impacts	X	X	X	X	
Preliminary Results of Metering/Billing		√	√	√	
Results of Billing Analysis (with 1 full year post)			X for 92 prog.	X for 93 prog.	X for 94 prog.
Persistence estimates (2+ years post)				X for 92	X for 92-93
Process Results					
Start-up Review	√	X			
Customer Satisfaction		X	X	X	
Detailed feedback on procedures, etc.		X	X	X	

Key: √ = Preliminary or Minor Results

X = Major Results

2.1 Questions to be Addressed by the Evaluation Plan

Assessing the performance of a demand-side management program requires a methodology that addresses all the aspects of the program and recognizes the interrelationship between the customers, trade allies, and the utility. The following section describes the questions that the evaluation strategy presented in this document is designed to answer.

2.1.1 Market Assessment Questions

Market assessments utilize information from surveys of random samples of all customers, program participants and non-participants. This information is used to determine customer market profiles, the market potential for energy efficient products and improvements, and whether systematic differences exist between participants and nonparticipants. Market assessments also assess barriers to customer acceptance. Market assessments differ by market sectors and technologies. They are tailored to the overall characteristics of residential and nonresidential markets, the trade allies supplying the equipment and key customer characteristics.

In this plan, the market assessment is in many cases incorporated as a component of the process evaluation. In addition, some of the results of the market assessment will be incorporated in the impact analysis.

The questions to be addressed by the market assessment relate to the following key issues:

- **Issue 1: Appraisal of Current Market Penetration**

  How does actual program performance compare to expected achievements and goals, by measure and market segment? What are specific areas of over- or underachievement? What is the remaining market potential that can also be projected.

- **Issue 2: Distribution of Participation by Market Subsector**

  What are participation rates of the retrofit/replacement market as compared to those of the new purchase market? What are participation rates by type of dwelling, type of business and other market segmentation criteria? Answers to those questions can help target future marketing efforts and help improve impact estimates. This activity requires coordination between impact and process evaluation efforts.

- **Issue 3: Determination of the Trade Ally Energy Efficient Product Inventory and Ordering Practices**

  What is the effect of the program on stocking and ordering practices for energy-efficient equipment? Resulting changes in inventory distribution can be factored into calculations of program cost-effectiveness. This activity requires coordination between impact and process evaluation efforts.

- **Issue 4: Determination of Status of Replaced Equipment**

  What is the disposition of replaced equipment? Replaced appliances and equipment can be tracked by surveys and site visits to see if they remain in the service territory. The extent of resale or reuse of old equipment can be estimated and factored into calculations for program cost-effectiveness. This activity requires coordination between impact and process evaluation efforts.

2.1.2 Process Evaluation Questions

The process evaluation will serve to qualitatively and quantitatively assess the performance of each program. It will reflect the customers' (participants and nonparticipants) perspectives, program staff views, and the trade ally.

The seven key issues generally applicable to all programs' process evaluations are discussed below.

- **Issue 1: Document Program History and Progress**

  The history of a program's development often contains important clues to understanding program design and implementation decisions. A review of history is also often a tool for guiding recommendations for program refinements. What are the significant factors in the program history and accomplishments to date? How have these factors affected the major program design decisions?

  These questions will be addressed through a review of relevant planning documents, program descriptions, annual and total goals and objectives, periodic reports of measures installed, costs incurred, interviews with program staff and trade allies. The sources for this review will include documents developed by Iowa-Illinois, the Iowa Utilities Board, the Office of the Consumer Advocate, and other relevant sources.

- **Issue 2: Program Targeting and Delivery**

  Although the Iowa-Illinois energy efficiency programs will be offered system-wide, there will be some variation in the types of targeting done, such as the lists of interested customers and matches between measures and customers' equipment/fuel combinations.

  Questions concerning target marketing and program delivery may include:

  1. How were contractors selected?

  2. How is contractor performance reviewed by Iowa-Illinois? What incentives or penalties are built into contractor performance review?

  3. How much operational discretion do contractors have to make decisions in the field? Does Iowa-Illinois exercise an appropriate level of control over the day-to-day activities of the contractors?

  4. What are the particular strengths of the installation process? What changes are needed to remedy any weaknesses? What actions by Iowa-Illinois and other involved parties will serve to support the program at a high level of quality?

  5. Are the materials and measure installations provided at a high level of quality? Are the rebated measures at an appropriate level of technology and reliability?

  6. Are all installations complete, and are completions achieved within a reasonable period of time?

  7. Are the contractors providing the installed materials at the agreed costs? What difficulties have they encountered

in assuring adequate supplies of necessary materials? What installation problems have been identified?

- **Issue 3: Identify Barriers to Participation**

  Iowa-Illinois has two principal concerns in marketing its energy efficiency programs in a diverse and widespread territory. Are customers who might benefit most from the program aware of it? Do they understand how the program will benefit them?

  A key evaluation question is whether the energy efficiency program is readily available to, and is being utilized by, the customers for whom it is principally intended. Are customers who are motivated to seek out the service able to easily gain access to it? Do customers who are solicited by Iowa-Illinois or its contractors understand what is being offered to them and can they accept the services on the basis on which the offer is made?

  Barriers among the rural communities are likely to be those of gaining access to the program. How easy is it for an interested customer to apply for and receive service? How long must customers wait from their expression of interest in the program to receiving services? Are contractors in sufficient contact with customers to minimize waiting time and missed appointments? How many visits are required to complete the process?

- **Issue 4: Evaluate General Program Satisfaction and Satisfaction with Specific Measures, and Estimate the Retention or Removal of Measures.**

  There are several aspects of customer satisfaction which are relevant to a complete understanding of customer perceptions of energy efficiency programs. Some of the questions to ascertain customer satisfaction include:

  1. How easy was it for customers to sign up for and receive the service?

  2. What were customers' expectations of what would be provided? How did those expectations match the reality?

3. How were the trade allies, contractors and Iowa-Illinois representatives perceived by customers? Did they appear when scheduled? Did they explain the services clearly? Did they do their work well and quickly?

4. Were the results what the customer expected? How did they differ?

5. Does the customer believe that Iowa-Illinois provided a valuable service? Do they believe they are directly benefiting from the service?

- **Issue 5: Assess Effectiveness of Marketing and Promotional Materials**

  The evaluation contractor will assess the effectiveness of the energy efficiency program's promotional and marketing materials in two ways. The primary measures of effectiveness of these materials will be the surveys of participants, refusers and the nonparticipant comparison groups. Using both "unassisted" and "assisted" responses, customers will be queried about how they heard about the program. Additional questions will explore how they were enrolled in the program; what they expected from reading, hearing or seeing promotions for the program; and what they learned from reading any educational or promotional materials. In addition to data gathered during customer surveys, the evaluation contractor staff will review energy efficiency brochures and other printed materials about energy conservation, buying energy efficient appliances and referrals to other Iowa-Illinois programs.

- **Issue 6: Identify Improvements for Program Design, Administration and Delivery Procedures.**

  To assess the operational effectiveness of the energy efficiency programs, the following questions will be answered:

1. Does the program reach its intended market?

   Who is the program serving? Are they the customers intended? What corrective measures can be taken to refocus on the intended targets of service?

2. Does the program deliver its services effectively?

3. Do the procedures, forms, and purchasing and reporting requirements support or inhibit effective program delivery?

What is the administrative organization of the program? Do the administrative systems support gathering, analyzing and reporting of the information in a timely, cost-effective manner?

Are there clearly delineated production and expenditure goals for Iowa-Illinois and for contractors? How are these goals expressed? How are they monitored? What mechanisms are there for corrective action if there are significant deficiencies in meeting production targets or in meeting budgetary goals? Are the problem-resolving mechanisms sufficient to deal with the problems encountered (or likely to be encountered)?

What changes in administrative organization or the organization of internal and external reporting and information will serve to optimize program delivery?

- **Issue 7: Qualitatively Assess Free-Riders, Free-Drivers and Snapback Effects.**

Free-riders are customers who benefit from conservation programs who would have undertaken the conservation measures without the benefit of the program. Free-drivers are customers who undertake measures on their own as a result of being influenced by the existence of a conservation program but do not participate in the utility program.

Snapback, or takeback, is an increase in customers' level of service from energy-using equipment following participation in a program. For example, customers may turn their heating thermostats up, experiencing increased comfort at approximately the same cost to them as before the measure installation, or may make use of rooms that were previously too cold or too expensive to heat at all.

Questions to be addressed in this area include:

- What actions would participants have taken without the program?
- What actions have nonparticipants taken as a result of the program?
- In what ways have participants changed their energy use behavior after participating?

These questions will be qualitatively addressed by asking customers about their energy-related decisions and behavior.

2.1.3 *Impact Evaluation Questions*

The purpose of the impact analysis is to quantify the energy and demand impacts of the program.

- **Issue 1: Energy and Demand Savings**

  The principal objective of the impact evaluations is to determine the summer peak demand reduction and annual energy savings by measure, measure category, and for each program as a whole. What were the amounts of the reductions actually achieved? What are the net effects of changes in weather and economic conditions?

  Iowa-Illinois plans to complete an annual impact analysis for each program in the field. The impact analysis plans presented in this document typically involve the use of multiple estimation methods. The use of multiple analyses will provide corroborating evidence to support the reasonableness of the estimated impact.

- **Issue 2: Free-Rider, Free-Driver, and Snapback Effects**

  What are the free-rider and free-driver effects? What are the net effects of these factors on program reductions? To what extent would customers who chose to participate have been expected to save? Would that be more or less than nonparticipants, even without the program? What degree of takeback (snapback) is evidenced by program participants?

- **Issue 3: Assessment of Measure Life and Persistence**

  What is the expected life of the installed equipment or structural changes? How long do the measures remain in place and operational? These questions will be addressed using annual household and business surveys and by site visits to a sub-sample of residences and businesses. Site visits will verify that installations actually did occur and have remained in place. Site visits will provide broadly representative estimates which are understood not to have high levels of precision due to limited sample sizes. Degradation of savings will also be addressed through annual updates of participant and nonparticipant cohorts with another year of billing histories.

- **Issue 4: Cost Effectiveness**

  Results of these analyses when combined with the results of the process evaluation provide the information necessary to calculate the cost-effectiveness of the program from the customer, utility and societal perspectives. Program net benefit to Iowa-Illinois will also be derived.

- **Issue 5: Validation and Refinement of the Assumptions Used to Derive Engineering Algorithms.**

  How accurate were the initial assumptions? What improvements can be made? Research results should help determine the difference between estimated reductions and savings and those achieved. This information will be used to refine the estimation methodology for planning and rebate calculations. A related benefit will be identification of systematic errors in data collected for engineering inputs and needs for other data.

- **Issue 6: Determination of power factor Impacts.**

  Most energy efficiency technologies reduce reactive load. Some energy efficiency technologies, however, may replace resistive load with new reactive load, for example, compact fluorescent lamps with extremely poor power factor that replace resistive incandescent lamps. Data gathered from participant applications, surveys, and site visits that describe technologies used for implementation can be used to calculate changes in customer power factor.

2.2 Summary of Evaluation Approaches

2.2.1 Process and Market Analysis Methods

The overall methodology for the process and market evaluations of Iowa-Illinois' DSM programs is similar across programs. The general approach for process evaluation is presented in this section. Tasks that pertain to a specific program are fully described in the individual plans. Table 2-1 shows the similarity across programs.

The evaluation will rely on both primary and secondary data sources. It is anticipated that the following data collection and analysis activities will be performed. Their integration into the overall evaluation plan's tasks are described more fully below.

- Review of program database and tracking system, including examination of selected program application forms.
- Interviews with utility program planning and implementation staff.
- Interviews with parties either involved with program design or significantly affected by the program.
- Participant surveys, including both the on-going "satisfaction" census of participants and annual surveys.
- Non-participant surveys.
- Trade ally surveys.
- Focus groups with trade allies and community groups (where necessary).

To the extent possible, data collection activities will be coordinated between programs. For example, data collection requirements for the residential high efficiency equipment rebate program may overlap with those of the residential new construction program.

Data collection activities will be phased in over a two-year period. This reflects different evaluation task requirements, and gradual

Table 2-1
Summary of Process Methodologies

A. Nonresidential Program

Methodology	Shopping List Rebates (NR-1)	Customized Rebates (NR-2)	Industrial Process Opt. (NR-3)	New Construction (NR-4)
Tracking System	X	X	X	X
Internal Evaluation				
• Review of Documentation and Procedures	X	X	X	X
• Review of Program Forms	X	X	X	X
• Review of Training Documents	X	X	X	X
• Interviews with Program Staff	X	X	X	X
• Trade Ally Interviews	X	X	X	
• Interviews with CAP Agencies				
• Focus Groups				
Financial Community, Property Mgrs., etc.				
Design Community				
Participant Group	X			X
Customer Surveys				
• Participant Surveys	X	X	X	X
• Nonparticipant Surveys	X	X	X	X

Table 2-1 (Continued)

B. Residential Programs

Methodology	Weatherization (R-1)	High-Efficiency Rebate (R-2)	Low-Income (R-3)	New Construction (R-6)
Tracking System	X	X	X	X
Internal Evaluation				
• Review of Documentation and Procedures	X	X	X	X
• Review of Program Forms	X	X	X	X
• Review of Training Documents	X	X	X	X
• Interviews with Program Staff	X	X	X	X
• Trade Ally Interviews	X	X		
• Interviews with CAP Agencies			X	
• Focus Groups				
Financial Community, Property Mgrs., etc.				
Design Community				
Participant Group	X	X		X
Customer Surveys				
• Participant Surveys	X	X		X
• Nonparticipant Surveys	X	X		X

Table 2-2
Relationship of Data Collection Activities to Process Evaluation Tasks

	Task 1 Program Development	Task 2 Program Targeting	Task 3 Barriers to Participation	Task 4 Program Satisfaction	Task 5 Review Mktg. Materials	Task 6 Admin. Procedures	Task 7 Freeridership & Snapback
Utility planning & implementation staff interviews	X	X	X		X	X	
Review program paperwork, completed applications & tracking system			X			X	
Review marketing materials		X			X		
Participant/non-participant mail/phone interviews		X	X	X	X		X
Staff interviews	X	X	X	X		X	X
Focus groups		X	X	X	X	X	X
Trade Ally Interviews/Surveys	X	X	X	X	X	X	X

Table 2-3
Process Evaluation Data Collection Schedule

	4th Quarter 1992	1st Quarter 1993	2nd Quarter 1993	3rd Quarter 1993	4th Quarter 1993	1st Quarter 1994	2nd Quarter 1994	3rd Quarter 1994
Utility planning & implementation staff interviews	X				X			
Review program paperwork, completed applications & tracking system	X				X			
Review marketing materials	X				X			
Participant/non-participant mail/phone surveys					X			
Focus groups		X				X		
Trade Allys/Interviews		X						

The first step in reviewing each program's history will be to examine current program offerings, marketing materials and internal utility memoranda addressing the program design process. Based on the findings from this review, the evaluation contractor will develop an interview guide to be used during the initial interviews with Iowa-Illinois program planning and implementation staff. The interviews performed during this task of the program evaluation will be semi-structured, allowing the interviewer to probe responses in more or less detail as required, and to follow up on areas of interest not previously considered.

Based on the initial interviews with Iowa-Illinois staff, the program contractor will provide for review and comments on the following:

- Revised interview guide.
- List of individuals and groups to be interviewed.

Based on comments from the Iowa-Illinois evaluation staff, the evaluation contractor will incorporate any changes to the interview guide and make any additions or deletions to the interview list.

The interviews for the non-utility parties will be made either in-person or by phone. For parties with a major contribution to, or a major stake in, the program, the interviews will be in-person (where practical, given time and budget constraints). Specific issues to be addressed to all groups during the interviews include, but are not limited to:

- Role and relationship of group to current program design and implementation.
- Specific concerns and how they were addressed and resolved during the program design phase.
- Program design options considered but rejected, and why.
- How program design decisions were made regarding the following:
  - Incentive levels
  - Measures included or excluded from program
  - Measure eligibility criteria
  - Marketing methods and materials

- Trade ally training
- Inspection and verification procedures
- Program staffing
- Program goals

Task 2: Examine Program Targeting and Delivery. As noted above, the design and implementation of a DSM program requires the input and active participation of a varied group of interested and affected parties. As a result, Iowa-Illinois' marketing efforts will be targeted to different groups and, ideally, the marketing message tailored to the intended audience. In this task the results from the participant and nonparticipant surveys, the program interviews from Task 1, trade ally focus groups, and the marketing material review in Task 5 will be synthesized to determine if the program has correctly identified the program's target markets and whether these efforts have been successful in generating program participation.

Review of the participant and nonparticipant surveys will allow an accurate demographic characterization of both populations, including:

- Type of home or building type.
- Urban/suburban/rural distribution.
- Home/facility size.

For participants, these data will be cross-tabulated by program measures installed. Additional analysis of these data will be performed to determine how the decision to participate was made and what the primary factors were that influenced the decision. Similarly, which party was the primary decision maker in deciding to participate? Finally, what program components and marketing approaches were most and least effective and why?

From the nonparticipant survey results, the analysis will determine reasons for not participating. These findings will be used both in this task and in the subsequent tasks: Barriers to Participation and Review of Program Marketing Materials. The nonparticipant analysis will address:

- Level of program familiarity
- Sources of information on program

- When problems arise, do the current administrative processes provide an adequate means for resolution?

The evaluation contractor will summarize the findings in this task and they will form the basis for Task 3, Identify Barriers to Participation, and Task 6, Recommended Program Changes.

Task 3: Identify Barriers to Participation. This task will review the results of prior data collection efforts and synthesize and summarize barriers, both real and perceived, to program participation. This program evaluation component will address and resolve the following questions regarding program participation:

- Are the program's benefits being clearly conveyed to the major target groups in this program, particularly to these groups and individuals who have been identified as primary decision makers regarding program participation?

- Is the current program design and associated training and marketing failing to address certain program benefits?

- What are the perceived costs associated with program participation? Can they be ameliorated or can they be addressed in such a manner so as to lessen their impact?

- Why, in the view of participants, do the program benefits outweigh the costs? Conversely, among nonparticipants, why do the costs outweigh the benefits, or are they not making a decision to participate because of inadequate program awareness?

Each of these issues will be addressed as a set of expanded questions, both open- and close-ended, during the data collection efforts. The program's benefits and costs will differ by respondent group.

For the trade allies, program benefits include the utility incentive. However, the perception of the magnitude of this benefit may be based significantly on how much of the incremental cost the incentive covers for any measure. Program costs include the time involved in dealing with program paperwork and inspections, and any effect these may have on the project schedule and cost.

Task 4: Determine Overall Program Satisfaction. Similar but different questionnaires will be administered to each participant for each program. Consistency among the survey instruments will be assured to allow direct comparison of results, where appropriate. In addition, each respondent group will be queried regarding the performance of all the other groups (i.e., trade ally, utility). Each group will be asked a similar set of questions pertaining to the utility program's staff performance and program operation:

- Clarity of application paperwork and ease in filling out the forms.
- Satisfaction with incentive levels.
- Timeliness of utility staff in processing applications, completing installation inspections and verifications, and issuing rebate checks.
- Accessibility of utility program personnel.
- Technical competence of utility staff.
- Adequacy of utility marketing of the program.

Task 5: Determine Effectiveness of Marketing Materials. This task will serve as a continuation of the marketing review undertaken in Task 2. The goal will be to qualitatively and quantitatively assess the effectiveness of the program's marketing materials. This determination will be based on the participant and nonparticipant surveys and the interviews and focus groups (where appropriate). These groups will be asked whether the program's benefits, as defined by Iowa-Illinois, are clearly stated and effectively conveyed by the marketing materials. Where different materials and different marketing approaches, e.g., direct mail, counter displays, etc., have been employed, the relative effectiveness of each will be determined. The evaluation contractor will also review the program's marketing budget to determine which marketing approach appears to be most cost-effective.

In addition to reviewing the responses to the marketing materials, the evaluation contractor will review all marketing materials and provide a written assessment of the graphics, program presentation and lost opportunities not addressed. Similarly, the marketing schedule will be reviewed to determine if information is being distributed in a manner consistent with construction timelines and decision-making criteria associated with new residential construction.

Task 6: Review Utility Program Management and Administration and Recommend Changes to Improve Program Efficiency. The evaluation contractor will review each step of the utility's administration of the energy efficiency programs. This review will also entail a thorough review of the program tracking database and program application and verification forms. The contractor will assess communication channels between utility staff and outside parties, as well as within the utility, paying particular attention to mechanisms that allow the program staff to pinpoint nascent problems and apply corrective action. Potential administrative bottlenecks will be identified and recommended changes suggested. Program staffing levels will be reviewed in relationship to the program's goals through the year 2000.

- Program tracking database/Program application forms
  - Do the program forms collect all the necessary information for this and subsequent process and impact evaluation?
  - Are useful program data collected on program forms, but not being entered into the database?
  - Are applications being processed in a timely manner?
  - How are blank fields being handled?
  - Is information being accurately transferred from the forms into the database?
- How easily can information in the database be accessed? What are its report generation capabilities?
- Are lines of authority clearly defined among the program staff?
- Are program responsibilities both defined and understood by the relevant individuals?

- Are program goals viewed as realistic and attainable given current and projected program resources?
- How are problems addressed by the program staff? Are mechanisms in place to insure that they are resolved quickly and effectively?
- What financial reporting requirements does the program have? Are costs being properly allocated and tracked?

Task 7: Assess Program Free-Ridership and Snapback Effects. The evaluation contractor will employ a variety of methodologies to measure program free-ridership and snapback effects. In general, two primary approaches will be used to measure these effects. The first will survey participants and nonparticipants on measure choices and occupant behaviors, while the second will compare pre- and post-program activities.

For program free-ridership, the participants will be asked what equipment they install and what equipment they specified prior to their program participation. Similar questions will be asked of non-participating customers.

2.2.3 Impact Analysis Methods

This section describes the method of impact analysis the evaluator will employ in the evaluation of Iowa-Illinois' residential, commercial, and industrial energy efficiency programs. Details of the analysis methodology will vary by program as conditions require. Variations in approach can be required by factors such as the timing of the analysis, lack of weather sensitivity in program measures, size of the program participant population and similar factors. If a program requires a substantially different approach, the evaluator will employ suitable alternative measurement methodologies.

Iowa-Illinois plans to complete an annual impact analysis for each program in the field. The impact analysis plans presented in this document typically involve the use of multiple estimation methods. The use of multiple analyses will provide corroborating evidence to support the reasonableness of the estimated impact.

The main goal of this impact analysis is to provide estimates of the annual energy and demand savings that have been produced by the

Iowa-Illinois energy efficiency programs. To do so, the most critical activity that must be undertaken from the outset of the DSM program is to document DSM implementation activities that were taken by, or for, Iowa-Illinois customers.

The impact evaluation emphasizes the use of the Iowa-Illinois program tracking system to provide preliminary estimates of gross energy savings per participant. Wherever practicable, billing analysis provides the final estimates of gross and net savings, based on changes observed in actual energy consumption. The billing analysis is conducted in two stages. The first stage determines gross savings for participants and nonparticipants. The second stage uses multivariate analysis techniques to isolate effects of various factors, and determine net program impacts. In some cases, billing analysis will be supplemented by special metering for more detail.

Tables 2-4 and 2-5 summarize the impact methodologies for residential and non-residential programs, respectively.

For the first program year, calendar year 1992, each DSM program impact evaluation will consist of the "implementation analysis" described below. Analysis of billing or other meter data cannot be performed for weather-sensitive technologies until at least one full cooling or heating season has passed following installation.

In the second and subsequent program years, billing analysis will be performed for all programs with sufficient participation. Billing analysis will also be carried forward from year to year so that empirical estimates of the persistence of savings can be reported.

Implementation Analysis: Engineering Estimates. In essence, an implementation analysis is an engineering analysis of installed measures. For this analysis, the program tracking system is used to aggregate expected energy savings, by end use, by measure, or by any other key variable that may provide useful information to the utility, such as geographic areas. The details of the implementation analysis are described in Section 2.4.1.

Engineering estimates will play several roles in the impact evaluations of Iowa-Illinois' programs. Perhaps the simplest is the projection of anticipated savings for programs before implementation. These projections, from the program plans, guided

Table 2-4
Summary of Impact Methodologies - Residential Programs

	R1 Weatherization	R2 High-Eff Eqt Rebate	R6 New Construction
Billing Analysis			
weather adjustment	yes	yes	yes
pre-post	yes	yes	no
engineering ests of energy savings	yes	yes	fundamental
peak impacts	small	moderate	moderate
Sample Design			
comparison group	matched NP direct install vs. insulation	matched NP	matched NP (small sample)
sample size	1200/year	1200/year	Census
Participation Modeling			
self-selection effect	small	moderate	moderate
logit P model	3-level	yes	yes
discrete choice tree	yes	yes	yes
Additional Data Collection			
On-site audits	100/year	100/year	100/year
Special metering operating hours	none	none	none
end-use interval	none	none	none

Table 2-5
Summary of Impact Methodologies - Non-Residential Programs

	NR1 Shopping List Rebate	NR2 Customized Rebate	NR3 Ind'l Optimization	NR4 New Construction
Billing Analysis				
weather adjustment	yes	yes	some	yes
pre-post	yes	yes	yes	no
engineering ests of energy savings	yes	fundamental	fundamental	fundamental
peak impacts	moderate	large	large	large
Sample Design				
comparison group	matched NP	None	None	matched NP (very small)
sample size	200	Census	Census	Census
Participation Modeling				
self-selection effect	moderate	large	large	large
logit P model	yes	yes	yes	yes
discrete choice tree	yes	no	no	yes
Additional Data Collection				
On-site audits	40	up to 10 each year	up to 10 each year	Census
Special metering operating hours	40 lighting	Lighting	Motors, major process eqt	Lighting
end-use interval	cooling heat pumps	HVAC	Motors, major process eqt	HVAC

decisions about the programs being developed, and in many cases can guide decisions about the programs as they are evaluated.

More important will be engineering estimates based on data for the current programs as actually implemented. For programs and measures that can be evaluated via billing analysis, the engineering estimates will provide preliminary estimates several months before the billing analysis results will be available. Thus, engineering methods will be used to provide initial estimates of 1992 program impacts before sufficient billing histories have accumulated. Subsequent comparison with these results will then indicate possible problems with the installations, with customer response to the measures installed, or with the data collected as a basis for the engineering estimates. Some of these possible problems will be investigated through the process evaluation. An additional role for the engineering estimates in conjunction with the billing analysis is to provide a basis for allocating total customer savings to individual measures installed.

For some types of programs or measurements, billing analysis cannot provide meaningful savings estimates. Billing analysis alone cannot measure demand (kW) impacts, except for demand-metered customers. For C&I programs that are tailored to individual customers, it may not be possible to establish a meaningful control group. In fact, if the DSM measures accompany other major changes at the facility, it is not possible to obtain even an unadjusted, gross savings estimate. Programs with small per-customer impacts or small participation levels may also be difficult to assess reliably with billing analysis alone.

For such situations, engineering analysis will be the primary basis for assessing savings. Wherever possible, the engineering estimates will be combined with information from billing histories or from special metering, to provide a stronger analytic foundation. In addition, experience gained from the billing analysis comparisons will help improve the engineering estimates even for cases where billing analysis is not possible.

Implementation analysis is driven by reported measures and their associated costs. In addition, the reliability of the program savings estimates is a function of the detail and building-type specifics that are reflected into the engineering equations or "algorithms." To insure that high quality estimates are available from the tracking system, XENERGY proposes to review the database and to enhance the algorithms on a case-by-case basis. XENERGY refers to this methodology as an "enhanced engineering analysis" approach to estimating energy savings.

Further, the engineering analysis will leverage the billing data analysis to improve future engineering-based estimates of program impacts. The billing analysis and customer survey data include achieved savings and customer characteristics, which can increase the accuracy of the benchmark data set used in the engineering analysis models.

Stage 1 Billing History Analysis. The first stage of savings estimation from billing data will rely only on these data. To statistically estimate program-related energy savings using billing histories, the evaluation will follow a pre-/post-installation, comparison group, quasi-experimental design framework. In this design, an account's annual energy use, calculated as the aggregation of monthly or periodic billing meter readings, is compared before and after energy efficiency measures have been installed. Long-term weather data and weather data for the analysis period are used to normalize pre- and post-installation consumption. To strengthen the research design, pre- and post-installation energy use for the participant group is compared to the energy use of a group of non-participants. The inclusion of nonparticipants in the research design offers a means of netting out the effects of general economic trends and free-ridership from changes in energy use that are due to program participation only. The non-participant adjustment also corrects for any systematic errors in the weather normalization procedures.

The advantages of these statistical analyses are clear:

- The resulting estimates reflect the savings that actually occurred, as observed in the customer's bill. They necessarily incorporate any snapback effects, actual usage patterns, and actual appliance characteristics.

- Because more economical billing data are used, rather than expensive metered data, larger samples of customers can be used in the analysis. Billing data can be obtained for customers for a period before and after participation in the program. Large samples allow for greater statistical precision in estimating savings.

For some programs, there is no pretreatment condition, and/or no valid comparison group. Special adjustment procedures are used to derive net program impacts in those cases. However, the Stage 1 billing analysis may be similar.

When valid pre/post periods and comparison groups can be defined, the primary limitation is that the savings from some measures might be too small relative to the customer's total consumption to be captured in the analysis of the customer's monthly bills. In most cases, however, the measures the customers install can be expected to cause a sufficient decrease in the customer's bill to be detected by the statistical analysis. Furthermore, it may be possible to identify relatively small savings by using the approach of entering the engineering estimates as explanatory variables in a Stage 2 model.

Stage 2 Analysis of Program-Related Energy Savings. The second stage of the billing analysis incorporates other customer information to refine the estimates of program performance based on billing histories alone. While the Stage 1 analysis offers preliminary estimates of gross and net energy savings, the reliability of the findings rests almost entirely on the assumption that participant and nonparticipant groups are, on average, similar in their energy use profiles and energy-related behaviors. To a limited extent, the sampling plan for an evaluation can attempt to assure comparability by matching participants and nonparticipants on critical variables. Nevertheless, there is wide room for differences to be present between groups.

To relax the "all things being equal" assumption, account-specific data will be collected via surveys. These surveys will be coordinated with the process evaluation portion of the study so that the objectives of both parts of the study can be achieved with a single data collection instrument. The focus of the survey for the impact evaluation will be on information that could be used to control for non-program changes that could have affected energy use during the study period. In particular, information on participation in other energy efficiency programs, independently-initiated energy conservation investments or improvements, and general changes in the building/equipment structures or functions, will be sought. Other household-or business-specific information, such as the square footage, size of household or number of occupants, and heating/hot water fuels will also be sought if they are not present in the program data tracking system.

If those customers who participate in the program are also the ones that have the most to gain from the program, net savings estimates can be inflated when projecting the results to future participants. In addition, in this case, the nonparticipants are unlike the participants, even without the program. As a result, the difference between the two groups is not an accurate measure of the program effect.

The survey data will be used to develop two types of regression models. The first models the decision to participate. The second relates energy use to customer characteristics. If necessary, based on analysis with the first participation model, a correction term developed from that model will be included in the second.

Sample Metering. As an additional strategy, metering will be installed in some of the programs to provide supportive data. Metered information is important for measuring the capacity, or load shape, impacts of the energy efficiency programs. The results from the metering will be used to enhance the engineering models, and will also be applied in the billing analysis.

The use of metering depends upon several different factors. Metering is warranted in situations where the benefits of the increased precision offered by this technique outweigh the costs. In general, this can only be the case where estimated savings are very high and/or penetration of measures is very great. These conditions can be met when a single site, such as in the case where a manufacturer is making a large investment in energy efficiency, or when a single end use, such as commercial lighting, is adopted by many customers.

The general strategy adopted in this plan is to use metering in cases where demand impacts are expected to be high, and where energy impacts are difficult to assess via billing analysis. These cases occur mainly in the non-residential programs. For these cases, run-time meters are recommended for lights and end-use interval meters for motors and HVAC equipment that is expected to have large impacts. In all cases, it is recommended that the metering be initiated prior to measure installation and continue afterward, to provide pre/post comparisons.

Due to the high costs associated with metering and the modest size of Iowa-Illinois' programs, it is not easy to decide in advance what the optimal number, duration, and sample plan should be for

PRISM diagnostics also provide a means of calculating the accuracy of peak-day savings estimate.

As for the annual energy savings, the PRISM-based estimate of peak day savings is for the whole unit, not for individual measures. The whole-unit peak savings can be allocated to individual measures on the basis of engineering models, using regression methods similar to those for the annual savings.

Peak Hour Electric Demand Impacts. Some of Iowa-Illinois programs are likely to have little impact on the annual peak demand. To avoid elaborate analysis of small effects, the evaluator will first conduct a preliminary assessment of the likely magnitude of demand impacts for each program measure. For programs or measures likely to have little impact on peak summer demand, simple estimates will be provided, based on scaling the energy impacts. For programs likely to have more substantial impact on peak summer demand, a more detailed analysis will be performed.

There are a number of models which could be used for this detailed analysis. The two best candidates are ESPRE™ developed by EPRI and CALRES™ developed by the California Energy Commission. Consideration must also be given to non-heating and cooling loads. Specifically, water heating, refrigeration and lighting end uses are subject to program measures. We will identify and justify the best models (or load shapes for non-HVAC end uses) appropriate to Iowa-Illinois. Some models may support a gas peak day analysis to be used along with the PRISM peak day analysis. The evaluator will select an appropriate model in the context of the project.

Regardless of model selection, the integration process will make use of the best available Iowa-Illinois data for calibration and the final billing analysis estimates will be used to constrain the peak impacts to be consistent with the annual energy impacts. This process has five steps.

<u>Step One</u>
Select a group of customers to be modeled. The group may be selected based upon specific customer characteristics or upon the availability of engineering model inputs. Consideration can also be given to the availability of metered data, either end use or whole building. This will be resolved separately for each program.

Step Two
Calibrate the Model. This may not be a statistical reconciliation depending upon data availability. However, the comparison, analysis and adjustment process will at a minimum be a validation against some level of Iowa-Illinois data to assure the reasonableness of the analysis. For non-HVAC end uses, this step may simplify into the collection of Iowa-Illinois end-use data being developed for other program evaluations. At this point the objective is to develop a representation of hourly end-use loads in the Iowa-Illinois service territory, for conditions in that territory.

Step Three
Simulate loads over the course of a year. This step will generate two outputs. First, peak day and annual energy data will be developed. Secondly, hourly profiles for the system peak day will be generated.

Step Four
Develop the relationship between peak day and annual impacts. The ratio of peak to annual energy from the simulation model provides the consistency link to estimated annual energy impacts. This ratio can be applied to each measure estimate (by end use) to generate the estimated peak day energy impacts. The net to gross adjustment can next be applied after the first year billing analysis.

For gas, the peak day analysis is complete. For electricity savings measure the peak hour is addressed in Step 5.

Step Five
Analyze the hourly change in the end-use demand. For each measure we can have the daily total impact. Some measures represent a constant savings while others are proportional to use. For example, water heater wrap impacts are a function of losses, not usage. Insulation saving air conditioning is a function of the pattern of cooling requirements. For constant savings measures any peak hour impact is estimated at one-twenty-fourth of the daily energy impact. For usage-proportional measures, the fraction of the utility's peak hour usage to peak day usage times the peak day measure impact is the appropriate demand impact.

Persistence and Degradation. Persistence will be explored through the participant surveys. Questions will be designed to assess whether each technology is functioning satisfactorily from the customer's perspective. Analysis of the responses will signal the need for additional investigation by Iowa-Illinois staff. Verification of installation and continuing operation will also be provided by on-site inspections. Ten case studies of each technology will be performed from a random sample of participants who have had the technologies for at least six months. The sample for each program year will be tracked by follow-up phone surveys in subsequent years through the end of measure life. Any problems so identified will be pursued by on-site visits by program or evaluation personnel where appropriate.

Degradation will be tracked by longitudinal studies of repeated billing analysis in subsequent years. This will address both behavioral (snapback and educational effects) and technology performance aspects on a net basis.

2.2.4 *Impact Evaluation Tasks*

A similar sequence of tasks will be used for all the impact evaluations. The tasks for the residential programs are summarized below. Details or exceptions for the individual programs are discussed in the sections on each program. The impact analysis for the Low-Income program is being conducted separately, as part of the statewide low-income program evaluation. For the non-residential programs, similar steps are followed, but the approach varies more across programs.

The following describes the tasks that will be undertaken for the impact analysis.

Task 1. Select Samples. The participant and nonparticipant samples will be drawn according to the guidelines described in Section 2.3.2. The proportions of customers in each category of billing completeness will be determined. For the participant sample, the distribution of participation dates will be determined. Tracking information will be linked to the billing records.

Task 2. Collect Survey Data. The participant and nonparticipant surveys will provide socioeconomic data including:

- Appliance saturation;
- Size of household;
- Appliance usage information;

- Fuel type data; and
- Occupancy patterns, etc.

This information will be used as inputs to the engineering estimates, as part of the implementation analysis. The survey will also collect data required for the process evaluation.

Task 3. Conduct On-site Audits. An on-site audit will be performed for a sample of participants. The sample will be stratified by dwelling or building type. The purpose of the audits will be to collect in-field data to verify the engineering assumptions. The audits will collect information on dwelling or building characteristics such as:

- Number of floors;
- Square footage;
- Insulation levels; and
- Infiltration rates, etc.

Task 4. Refine Engineering Models. The on-site data will be used to refine the input assumptions used in the engineering models. Depending upon the final design of the program, these data will be incorporated into either engineering algorithms or a thermal simulation model.

Task 5. Estimate Savings Using Engineering Models and Conduct Implementation Analysis. The engineering models will estimate the savings attributable to various DSM measures and combination of measures offered under the energy efficiency programs. When the results from the process evaluation are available, the gross savings produced by the engineering models will be adjusted for free ridership and snapback. This will be conducted in September 1992 and in March 1993. The first implementation analysis will serve as a preliminary analysis of early program participation. The latter analysis will be the basis for deriving engineering estimates for the full program year for Iowa-Illinois' regulatory filing.

This task includes a comprehensive review of the data collected in Iowa-Illinois' tracking system. The evaluation team will review the quality and completeness of the database. Where necessary, recommendations will be made to improve the data collection process. The analysis of the data will characterize the type of customers and dwellings participating in the programs. At a minimum it will include:

- The number of participants;
- A breakdown of participants by dwelling type;
- The quantity and type of measures installed;
- The efficiency levels of measures installed; and
- Square footage of dwelling.

Task 6. Conduct Billing Analysis (Stage 1). As described in Section 1, analysis of billing data will be used to quantify the savings impacts from the programs. The analysis is a two-stage approach. Stage one involves the comparison of weather normalized data. Stage two utilizes multivariate regression models and statistically adjusted engineering models to estimate energy savings.

Task 7. Conduct Multivariate Regression Analysis (Stage 2). This task consists of three main steps, described in detail in Section 2.4.3.

- Step 1. Develop the participation model. This step identifies those factors that are associated with a greater likelihood of participating in the program. The model estimates the likelihood of participation as a function of the factors collected in the participant and nonparticipant surveys. If none of these factors are found to have a significant relationship to the decision to participate, no self-selection correction is made in the next step.
- Step 2. Develop the econometric energy model. This model relates the energy (or demand) savings (or consumption level) to the physical and socioeconomic factors known for participants and nonparticipants. Included among these factors is the propensity to participate estimated from Step 1.
- Step 3. Use the model from Step 2 to estimate the effect of free riders, and the net effect of the program.

Task 8. Estimate the Demand Impacts. Demand impacts will be estimated according to the procedures described in Section 2.2.3. Iowa-Illinois' current load research data will be used for the Weatherization and High-Efficiency Equipment Rebate programs. For the New Construction program, load data will be collected from a sample of participants and nonparticipants.

2.3 Data Collection

Each component of the evaluation will rely on data from several sources. Each data collection effort is designed to serve multiple purposes. This approach helps to provide consistency across evaluation components, as well as making more efficient use of project resources. In addition, minimizing the burden on respondents will improve survey response rates, resulting in more accurate results.

2.3.1 Data Sources and Timing

Data will be collected from Iowa-Illinois and its implementation contractor, from the utility's customers, and from outside sources. The data collections from each of these sources are described below. The general evaluation areas that each data collection activity is applied to are summarized in Table 2-6. The schedule of these activities is indicated in Table 2-7.

- **Data from the Utility**

  Tracking System. Information from the tracking system will be used directly in the process and impact analyses. In addition, the tracking system will provide the basis for linking information from other sources. As a result, tracking system information will be accessed by the evaluation team on an ongoing basis throughout the course of the project. In the early stages, the tracking system data items and organization will be reviewed, and suggestions made for improvements.

  Billing Records. For most programs, billing records will be collected for a sample of participants and a sample of nonparticipants. The records will span a period corresponding to one year before and one year after participation. Depending on the specific program, records may be taken for either all participants or a sample. For the

Table 2-6
Data Collection for Evaluation Components

	Market Assessment	Process Evaluation	Implementation Analysis	Billing Analysis Stage 1	Billing Analysis Stage 2
FROM THE UTILITY					
Tracking System		X	*	X	X
Billing Records			X	*	X
Program Staff Interviews	X	*			
FROM CUSTOMERS					
Participant Follow-up Survey	X	X			X
Participant/Nonparticipant Characterization Survey	X	X			X
On-site Audits		X	X		
Special Metering					
- Case Studies				X	
- Subsectors				X	X
FROM OTHER SOURCES					
Trade Ally Surveys/Interviews	*	X			
Published Data				X	X

X = Data Element Used for this Evaluation Component

* = Key Data Element for this Evaluation Component

new construction programs no pre-program data will be taken. For the tailored nonresidential programs, which will be evaluated by cases studies, no nonparticipant data will be taken. These records will be assembled in the third quarter of 1993 for preliminary assessment of early program participants, and in the third quarter of 1994 to capture a full (or nearly full) year of post-participation data for the entire 1992 program year.

Program Staff Interviews. Interviews with program staff from the utility and its implementation contractors will be a key element of the process analysis. These interviews will be conducted during the fourth quarter of 1992 and again in the fourth quarter of 1993.

- Data from Customers

Participant Short-Term Follow-up Survey. One of the major objectives of the process evaluation is to assess factors influencing program participation. Accurate information on customers' program awareness and decision-making criteria is difficult to obtain several months after the fact. A brief follow-up survey will be administered to program participants within a few weeks of participation. This survey will not obtain detailed customer characteristics, but will focus on those decision-related factors that might be difficult to recall by the time of the main participant/nonparticipant survey.

Participant/nonparticipant survey. A survey will be administered to a sample of participants and nonparticipants at the end of the first program year. Results from this survey will be used by both the process and impact evaluations. The information collected will include characteristics of the premise, its occupants, and its operation.

On-site Audits. For each of the programs, on-site audits will be conducted of some or all of the participants. The audit will determine if the equipment is in place, in use, and operating as intended. Input engineering assumptions, such as operating hours and floor space, will be verified. Some follow-up information on customer satisfaction may also be collected at this time. This information will be used on an anecdotal basis to refine the implementation analysis, to explain and support results of the impact evaluation, and to provide further information for the process evaluation.

Most of the on-site audits will be performed in the latter half of 1993. For customers whose main DSM actions related to cooling (or lighting) loads, the on-site audits will be scheduled for the warmer weather. For actions targeted more at heating loads, the audits will take place during the heating season. In addition, for those customers where special meters are installed for the evaluation, on-site audits will be performed at the time of the meter installations.

Special Metering. Special metering will be performed primarily for the nonresidential programs. Metering will be done for all customers who are subjects of case studies for the tailored programs. In addition, some load research data will be collected for the new construction evaluations. The case-study metering will be performed on an ongoing basis, since before and after data are to be collected at the time of program participation. Metering to obtain a profile of a customer subsector will be conducted in the third quarters of 1993 and 1994.

- **Data from Other Sources**

Trade Ally Surveys and Interviews. Sales data from trade allies will be used as input to the market assessment, as background on participation decisions for the process evaluation, and to provide baseline efficiency levels for the impact analysis. Information will be collected from trade allies on a semiannual basis, in the first and third quarters of 1993 and 1994.

Published Data. For the billing analysis, weather data will be obtained for local weather stations from the National Oceanic and Atmospheric Administration. For the multivariate analysis performed as the second stage of the billing analysis, local and national economic indicators will be used. These external data will be obtained from the most recent sources available at the time the analysis is undertaken.

2.3.2 Sample Design

The impact evaluation relies on energy and customer data. For most of the programs to be evaluated, these data are collected for a comparison group of nonparticipants as well as for the participants. The process evaluation similarly relies on customer data from surveys and from the tracking system. As indicated above, the data collection and analysis for these two components of the evaluation will be coordinated. Accordingly, the sample design must address the needs of both components.

The general approach of this evaluation design is to use a comparison group of nonparticipants wherever possible. In each sector, a master nonparticipant sample is drawn to serve the needs of all applicable programs. Thus, there are four overall groups to be considered: residential participants and nonparticipants, and nonresidential participants and nonparticipants. A similar design is used within each of these groups.

The sample frame is taken from the customer billing records for each sector (residential and nonresidential). First, records for all participants (in any of the programs for that sector) are compiled. The remaining records constitute the nonparticipant frame. Each group is then divided according to the number of years the customer has been in service, and whether it is a new construction case. Based on the years of service, customers are divided into the following categories:

- Long-term: Two or more years of service
- Recent movers: One to two years in service in previously existing building
- New construction: built within the program year
- Too recent: less than one year in service. These customers will be excluded from the analysis.

Iowa-Illinois uses meters mounted either on the exterior of the building or in accessible internal locations. Meters are typically read monthly and therefore most billing histories are expected to be complete. The most likely reason for incomplete billing histories would be frequent shut-offs. This would be a problem primarily for analysis of the low-income program. As discussed in Section 6, the low-income program evaluation will only address process issues; estimated impacts will be taken from the statewide evaluation. For this particular process evaluation, customers with incomplete billing histories will be included in the sample.

Each of these groups may be further divided according to utility district or neighborhood, energy or demand level, and SIC. The evaluator will allocate the sample among the different strata according to the distribution in the customer population. Certain groups may be sampled at higher rates to ensure adequate analytic cases for particular subgroups of interest. For example, certain neighborhoods might be targeted if they are believed to have high proportions of electric water heaters. The disproportionate sampling will be corrected for in the analysis by using different weights, to avoid biased results.

A survey will be administered to each sampled participant and nonparticipant. As described above, the survey collects information on the building structural and equipment characteristics; fuels used for heating, cooling, and hot water; measures installed during the program year; and participation in other programs. Additional information is collected from participants to aid in the engineering and customer satisfaction analysis. For those items to be compared between participants and nonparticipants, the questions will be asked as similarly as possible.

The long-term customers are the basis for the pre-post comparisons for participants and nonparticipants. For new construction, the nonparticipants are compared directly to participants, without a pre/post difference. The recent movers can be analyzed by methods similar to those for new construction.

Sample Design for Tailored Programs

The Customized Rebate Program and the Industrial Process Optimization Program are tailored to individual commercial and industrial customers. Because the needs and actions are likely to be very specific to the individual customers, no comparison group will be used. These programs will be analyzed by case studies of participating customers. Depending on the numbers of participants, all participants may be studied. If the number participating is larger than anticipated, a sample may be taken. The sample will be taken using similar stratification to the other nonresidential programs.

Sample Sizes

Sample sizes are set according to a combination of criteria. The most straightforward is the number of sample cases required to ensure estimates with a given level of statistical precision. Another is to have sufficient cases to provide meaningful estimates for various subgroups of interest. A related goal is to have a large enough number that relatively rare customer types will still be likely to be found in the sample.

The precision criteria for this evaluation has been set at 10 percent precision for the mean consumption (or demand) level (pre- or post-program) at 90 percent confidence. If the coefficient of variation (ratio of standard deviation to mean) for the population is 0.50, a sample size of 68 would be sufficient to meet the 90/10 criterion.

In order to meet additional evaluation objectives, the sample sizes proposed for the residential and nonresidential master samples are considerably larger than required to meet the minimum precision criteria. These objectives are:

- To allow accurate estimates of savings, computed as the difference between pre- and post-program levels, and/or the difference between participants and nonparticipants. This is particularly important when the mean savings are likely to be small.
- To allow accurate estimates for customer subgroups, defined by utility district or measures installed.
- To ensure an adequate matched comparison group for certain customer characteristics. In particular, to ensure an adequate comparison group of equipment purchasers for the rebate programs.
- To provide a broad customer profile, especially for nonresidential customers.

The sample size proposed for the residential master sample is 1200. This sample size will allow accurate estimates for some subgroups. In addition, a fairly large master sample is required to ensure meaningful sample sizes for special comparison groups. For the rebate programs, in particular, it is necessary to find nonparticipating customers who have replaced major equipment.

Assuming a 12-year life for a water heater, 100 water heater replacements would be expected in the nonparticipant sample for each program year. Of these, fewer than 10 would be electric water heaters, unless a strategy is adopted to identify more.

2.4 Details on Impact Evaluation Methods

2.4.1 Implementation Analysis

As an initial activity for the program's impact evaluation, Iowa-Illinois' program tracking systems will be used to perform an "implementation analysis." XENERGY will check the tracking system to assure that the data are accurate, comprehensive and uniform, and that a consistent set of decision rules and algorithms are used to generate program information. On a quarterly basis, XENERGY will generate program summary statistics in formats that are concise and understandable.

The end product of the implementation analysis will be statistics that include:

- the number of "participants";
- the total number of installed measures, by end use;
- the average number of installed measures per account per end use (some C/I customers have multiple accounts);
- total and average engineering-estimated energy/demand savings for measures implemented, by end use;
- total and average costs (either estimated or actual) for measures implemented, by end use (both including and excluding zero-cost measures);
- estimated simple payback for measures installed.

Though straightforward in principle, devising a consistent accounting system for participants, customers, accounts, and meters has some complications. Some customers may have participated in the same program more than once, for different measure installations, so would appear two or more times in participant lists.

A customer also may have multiple accounts and/or multiple meters. The simplest accounting system would treat each participation as a separate "participant." However, the billing analysis is for the customer, and would include the effects of all measures installed. Care must also be taken to ensure that measures installed in a multiple-meter facility are associated with the correct meters. A related complication is the allocation of program costs to individual customers and measures. XENERGY will work with Iowa-Illinois to agree on a scheme that is clear and consistent with the overall evaluation goals.

In addition to these program summary statistics, the evaluator will undertake a periodic review of program data collection and processing quality. This activity is critical for assuring that accurate, consistent, relevant and complete data are available for subsequent program evaluations. The quality control checks that will be instituted will include examination of the tracking systems for:

- variables with a high proportion of missing values;
- for discrete variables, an inordinately high proportion of a single value where not determined by common physical attributes;
- for continuous variables, high and low outliers, and the proportion of zero values;
- for continuous variables, skewed distribution or distributions with coefficients of variation (computed as the standard deviation of a variable divided by its mean) greater than 0.5.

Once the program tracking system is validated and analyzed at the aggregate level, XENERGY will provide program statistics at the specific levels of aggregation and disaggregation, e.g., by geographic area, energy use levels, or customer profile, desired by Iowa-Illinois.

2.4.2 Billing Analysis: Stage 1

Meter-reading Cleaning. To initiate the statistical analysis of energy savings, approximately two or three years of billing history data must be collected for all program participants. Pre-, mid- and post-installation periods for each of these accounts will be determined based on installation start and end dates reported for each customer on Iowa-Illinois' program tracking database. Before the billing histories can be used to study energy savings, they must be carefully screened and processed. The billing history raw data screening included the following procedures:

- Billing accounts that contain missing or estimated data points, or that contained zero readings are flagged. Estimated meter readings are combined with the next actual reading to create a longer period bounded by actual readings.

- Billing histories with meter reading dates that are discontinuous, that are out of sequence, that do not match the number of days reported in the cycle, or that contain many fewer or many more days in the cycle than are typical, are flagged. These cases are reviewed to determine if the data should be assumed correct as reported.

- Billing histories offering less than a specified number of meter readings in either the pre- or post-installation periods are flagged.

On the basis of this screening, the study cases are divided into three groups:

(1) cases with too few actual reads to fit the model;

(2) cases with sufficient observations, but flagged for serious irregularities;

(3) cases with adequate, clean billing histories.

Model Specifications for Billing Analysis. After screening the billing histories, weather-adjusted annual consumption is estimated for each account. Weather adjustment and annualization is necessary to ensure that the comparisons of energy use between years, and between participant and non-participant groups, span the same number of days and the same temperatures. In addition, weather-adjustment involves forecasting annual energy consumption using long-run, average temperatures. This allows the results of the analysis to be viewed as valid for the lives of the installed measures, rather than just for the years that make up the study periods.

Weather-adjusted annual consumption is forecast for each account by estimating separate multiple linear regression models for each account in the pre- and post-installation periods. One program that performs this function is PRISM, the Princeton Scorekeeping Method. The functional form of the time-series model is:

$$kWht = a1 + a2(HDDt(T)) + et$$

where "kWht" is the average daily consumption for a given account in period t; "HDD" is the average heating degree days at a reference temperature T for a given account in period t; and "T" is a reference temperature that is determined by the objective function of minimizing the standard error of the model. The term "et" is the residual error for the time period t.

To annualize and weather-normalize, a value representing the long-run (e.g., 20 year) average daily heating degree days, LRHDD, is inserted into the model. To calculate normalized annual consumption, or NAC, the formula is:

$$NAC = a1(365) + a2(LRHDD(T))(365)$$

For this study, XENERGY proposes to use PRISM (the Princeton Scorekeeping Method). However, XENERGY has found that in a number of circumstances, PRISM cannot, or should not, be employed. For example, if less than eight post-installation months are available per account, or a full heating season is not available, PRISM estimates are likely to be unreliable. In these situations, other billing analysis models will also be unreliable for individual cases. However, it may still be possible to obtain meaningful results for the average or aggregate impact.

Another concern with using PRISM is that a cooling adjustment as well as a heating adjustment may be appropriate. XENERGY will investigate the possibility of using the Heating-Cooling version of PRISM currently under development. Alternatively, the analysis may rely on a more general time-series model, for example, of the form:

$$kWht = a1 + a2\ HDD(T_1) + a3\ CDD(T_2) + a4(MID) + a5(POST) + et$$

where "kWht" is the total consumption for a given account in period t; "$HDD(T_1)$" is the total heating degree days at a base of T_1 (e.g., 55 degrees Fahrenheit) for a given account in period t; and, "$CDD(T_2)$" is the total cooling degree hours at a base of T_2 (e.g., 60 degrees Fahrenheit) for a given account in period t. The variables "MID" and

"POST" are indicator variables; MID designates the period(s) during which the installation was taking place, and POST designates the post-installation meter reading periods.

In the model just specified, "a1" is a constant (the regression intercept) that is interpretable as non-weather sensitive load or baseload per period; "a2" and "a3" are coefficients providing estimates of how a marginal change in temperature affects electricity consumption; and "a4" and "a5" offer estimates of the change in energy use per meter-reading period during different phases of the installation.

Since no conclusions can be drawn about energy use in the mid-installation period, the term a_4 MID is of not interest for the present study; it remains in the model simply to avoid the loss of meter-reading observations. However, because energy use patterns can be erratic in the midst of the installation period, including observations from this period could introduce additional noise into all the estimates of interest. If so, the MID term and corresponding observations will be deleted from the model. This determination will be made on the basis of a statistical test of the net improvement or degradation of estimation accuracy for the other estimates.

The symbol "et" completes the model by representing unexplained variation, that is, the kWh per period that remains unaccounted for by the model's estimates. As with the PRISM model, the error term is assumed to have a normal distribution, with a mean of zero, and a constant variance.

Both the PRISM model and the slightly expanded model above are very simple physical models of energy consumption. Although a more complicated model could be specified, incorporating additional physical drivers, experience has shown the PRISM model to be extremely effective in spite of its simplicity. The effects of omitted variables that vary seasonally, such as wind and sun, will be incorporated in the estimated coefficients for HDD (and CDD), but empirical studies have shown these effects to be relatively small. The omission of variables that have no seasonal component will have no distorting effect on the estimated coefficients.

The expanded model, which fits pre- post- and mid-installation periods simultaneously, assumes that the building's responsiveness to temperature is not affected by the program-related changes. Depending on the nature of the installation, this may or may not be a good assumption. The assumption can easily be tested, as with interaction terms, and the specification modified accordingly.

Although it is not unusual for many time series models to suffer from first-order auto-correlation, for this application, as for PRISM, the correction for auto-correlation is not desirable. One reason for this is because first-order correction results in the loss of the first meter-reading period; given the limited number of meter-reading periods in each account, this loss can cause more problems than the auto-correlation itself. A second and perhaps more important reason is that auto-correlation does not affect the consistency or unbiasedness of the model estimates. Rather, it results in estimates for individual buildings that appear to be more accurate (have smaller standard errors) than they actually are. However, the accuracy of the mean savings is still correctly determined from the spread of estimated savings across the sample. Since the purpose of this application is to derive point forecasts of annual energy use for the program as a whole, maintaining an unbiased model is far more critical than maintaining a model with accurate estimates of error for individual customers.

To construct weather-adjusted annual consumption in the pre- and post-installation periods for each account, values representing 20 year, long-run heating and cooling degree days (LRHDD and LRCDD) for a typical, 30 day meter reading period, can be inserted into the estimated model. The formulas for calculating pre- and post-installation, weather-adjusted, annual consumption are:

$$Pre\_kWh = (a1 + a2(LRHDD) + a3(LRCDD)) * 12$$

$$Post\_kWh = (a1 + a2(LRHDD) + a3(LRCDD) + a5) * 12.$$

The model fits will be screened for outliers, both with respect to the parameters estimated for each year and with respect to the savings estimates. To the extent practical, these cases will be reviewed for possible data errors, or for major changes unrelated to the DSM program. Additional flags will be set to indicate possibly unreliable fits. These and the original data screening flags will be used as selection and modeling criteria in subsequent analysis.

Analysis of Aggregate Savings. The customer-specific analysis of energy savings involves subtracting weather-adjusted, annualized energy use in the pre-installation period from energy use in the post installation period. This change is computed for each participant and each non-participant in the study group. To determine overall or average program effects, the average savings is computed across all participants and across all non-participants.

When examining the changes for participants only, the difference in energy use is often referred to as "gross savings." When the mean change in energy use of the participant sample is subtracted from the mean change of the nonparticipants, the difference is referred to as "net savings."

The distinction between gross and net savings is critical to the study of program-related impacts. Gross savings are simply the observed differences between pre- and post-installation use. By itself, this difference is a relatively unreliable indicator of program effect. This is partly because of the long-term trends in energy use in the economy that are caused by changes in technologies, changes in production mix, evolving tastes, and so on. Until the 1980s, this trend was towards increasing energy use; since then, the long-term trend has been towards stable or decreasing energy use. If these secular trends, which are embedded in the year-to-year differences in energy use, are not taken into account, energy use changes may be falsely attributed to program intervention.

Changes in the relative prices of fuels and other production inputs, changes in regional incomes, aggregate demand, and other cyclical economic trends, may also affect the levels of energy use of commercial accounts from one year to the next. For example, were a business to be in an expansion phase where it was adding equipment and employees, energy use may increase for this business' account in the post-installation period, despite the increased energy efficiency of its equipment. Conversely, in a business contraction, energy use may decrease by far more than might be expected by the newly-increased energy efficient equipment.

Changes may also occur as a result of unusual weather patterns, not accounted for by the weather normalization procedure. For example, in a mild winter, some customers might not put their storm windows on, or might keep the house warmer than usual. In a particularly windy winter, customers might require higher interior temperatures to be comfortable. The weather normalization procedures do not adjust for such behavioral responses. These procedures account only for differences in temperature, assuming the physical and operational characteristics of the house or building are constant.

The non-participant comparison group is designed to control for such secular and cyclical changes in energy use. Ideally, the non-participant study group possesses the same energy use characteristics on average as the program participants. The two groups are therefore similarly affected by outside factors. Subtracting the nonparticipant savings from the participant savings therefore eliminates the effects of nonprogram factors from the participant change.

Using a nonparticipant group invokes the broadest of all assumptions, the "all things being equal" qualification. This implies that the participant and nonparticipant groups consist of energy users that have the same makeup vis-a-vis types of businesses, types of buildings and equipment, and other energy use determinants. Along the same lines, it also implies that the groups share similar responses to exogenous business trends and random shocks, and that their free-market rate of adoption of energy efficient equipment levels are similar, save for program participation.

If this assumption holds, the nonparticipant group adjusts for free riders, the customers who would have taken action at their own expense but had the program pay for it instead. Since, by assumption, the nonparticipant group faces similar investment opportunities, the energy use of the nonparticipants incorporates the savings that participants would have enjoyed, at their own expense, without program participation. Taking the difference in the changes between the two groups should thus result in savings that are net of free-ridership, too.

The potential weakness of this approach is the validity of the assumption that non-participants represent how the participant group would have behaved in the absence of the program. This issue is addressed by the analysis of self-selection bias, in Stage 2.

2.4.3 Stage 2 Analysis of Program-Related Energy Savings

The second stage of analysis using billing data involves multivariate analysis to determine net program effect. This analysis stage fits two types of models. One models program participation. The second uses those results in a model of energy consumption.

The first model is a discrete choice (logit) model that tests for program self-selection, or the propensity of participants to

The non-participant comparison group is designed to control for such secular and cyclical changes in energy use. Ideally, the non-participant study group possesses the same energy use characteristics on average as the program participants. The two groups are therefore similarly affected by outside factors. Subtracting the nonparticipant savings from the participant savings therefore eliminates the effects of nonprogram factors from the participant change.

Using a nonparticipant group invokes the broadest of all assumptions, the "all things being equal" qualification. This implies that the participant and nonparticipant groups consist of energy users that have the same makeup vis-a-vis types of businesses, types of buildings and equipment, and other energy use determinants. Along the same lines, it also implies that the groups share similar responses to exogenous business trends and random shocks, and that their free-market rate of adoption of energy efficient equipment levels are similar, save for program participation.

If this assumption holds, the nonparticipant group adjusts for free riders, the customers who would have taken action at their own expense but had the program pay for it instead. Since, by assumption, the nonparticipant group faces similar investment opportunities, the energy use of the nonparticipants incorporates the savings that participants would have enjoyed, at their own expense, without program participation. Taking the difference in the changes between the two groups should thus result in savings that are net of free-ridership, too.

The potential weakness of this approach is the validity of the assumption that non-participants represent how the participant group would have behaved in the absence of the program. This issue is addressed by the analysis of self-selection bias, in Stage 2.

2.4.3 Stage 2 Analysis of Program-Related Energy Savings

The second stage of analysis using billing data involves multivariate analysis to determine net program effect. This analysis stage fits two types of models. One models program participation. The second uses those results in a model of energy consumption.

The first model is a discrete choice (logit) model that tests for program self-selection, or the propensity of participants to systematically differ from nonparticipants. The second is a cross-sectional model of energy use, that also uses the survey information to control for differences among participants and nonparticipants.

Self-Selection Model. To examine the issue of self-selection, XENERGY will develop a discrete choice model known as "logit" to control for self-selection bias. The model for this participation decision process is formalized by the equation:

$$Yi = B'Xi + Ui$$

where Yi is the unobserved or undocumented non-random motivation leading to building owner/manager i participating in the program, B' is a vector of unknown parameters, Xi is a vector of explanatory variables related to participation, and Ui is an error term. Though Yi is unobserved, we can distinguish the customers who participated; hence, Yi = 1 if a customer is a participant, and Yi = 0 otherwise.

Based on the binomial formulation, the probability of participation can be viewed as:

$$Prob(Y=1) = Prob(Ui) > -B'Xi = 1 - F(-B'Xi)$$

where F is the cumulative distribution function for u. Assuming that this function is bounded by 0 and 1 and is flatter at the extremes than it is in the middle, a logistic cumulative distribution for Ui can be specified, giving rise to a "logit" model. This model is formalized as:

$$F(-B'Xi) = 1 / (1 + exp(B'Xi))$$

or,

$$1 - F(-B'Xi) = exp(B'Xi) / (1 + exp(B'Xi))$$

Many variables related to the characteristics of buildings and building managers might be expected to influence the probability of participation, such as tenure (own or rent), energy price, energy-using systems, and so on. These variables will thus be part of the model specification. However, because other characteristics may yet be undiscovered, a systematic specification search will be conducted in which plausible independent variables will be tested in the model.

Once estimated, the decision model can be used to generate a "self-selection correction factor," C, that will be incorporated into the end use i; and V is a vector of variables related to economic and other exogenous conditions that could influence energy consumption during a given time period. Likewise,

$$Eo = fo(Mo, Do, V)$$

represents the building energy demand function for the set of unspecified end uses. As is evident, the key feature of the conditional demand model is that it allows observed total building consumption for a given fuel to be allocated among end uses, despite the fact that end-use consumption itself remains unobserved.

Using a linear form for the function fi, the demand equation for end use i would be written as

$$Ei = Mi'b_{Mi} + Di'b_{Di} + Vi'b_{Vi}.$$

Using this general framework to specify a multiple linear regression model, ordinary least squares can be applied to the data to estimate all the model parameters at the same time. Once estimated, the single, total conditional demand equation can be decomposed into individual equations. While each end-use function may share one or more variables with other functions, each end-use function will also have one or more variables that are unique.

Depending on the purposes of the individual program evaluations, the available sample sizes and the available data, it may be possible for XENERGY to save time and effort by developing a limited conditional demand model in which end uses will be represented by individual dummy variables indicating the presence or absence of specific equipment or building characteristics. More complex functions would generally provide more accurate end-use allocations for each customer. However, the dummy variable approach, with sparing use of additional more complicated terms, may be adequate to estimate average effects. Again, since the focus of the impact analyses is net program-related savings rather than end-use forecasts, some customer-level accuracy may be worth sacrificing in the interests of cost savings.

The model may be specified as an annual-level equation such that:

$$E_{i,t+1} = b_E E_{i,t} + b_M'Mi + b_D'Di + bp'Pi,$$

where $E_{i,t+1}$ represents energy consumption for the i'th building in the post-installation period t+1; $E_{i,t}$ represents energy consumption for the i'th building in the pre-installation period; Mi represents a vector of variables related to equipment ownership; Di represents behavioral and exogenous variables that may vary across households; Pi is the program impact variable that is coded 1 for participants and 0 for non-participants; and bj is a vector of coefficients for the corresponding independent variables, for j=E,M,D,P.

A simplified functional form for the model might take post-installation consumption (E) to be a function of the presence of electric space heating (SH) and electric hot water heating (WH), the number of people in the household (NofP), the size of the dwelling (SIZE), program participation (P), and an error term, e. Formally,

$$E = a1 + a2(SH) + a3(WH) + a4(NofP) + a5(SIZE) + a6(P) + e$$

where a1 is the intercept and a2 through a6 are parameter estimates. The average change in building electric consumption due to program participation is thus a6. Alternatively, the engineering priors can be used in the model to test the accuracy of the initial estimates of savings. In doing so, the total engineering estimate of energy savings due to the measures installed (MEA) for each household might be inserted into this specification in the place of P such that:

$$E = a1 + ... + a6(MEA) + e$$

in which case a6 is used to measure the degree to which the engineering estimate of program savings vary from the actual change in use. For this variable, the null hypothesis is that a6 is approximately 1.

3.1 Introduction

The proposed budgets for monitoring and evaluating Iowa-Illinois' eight energy efficiency programs are the result of considering many factors. Detailed plans for evaluating each program include specific activities for several years. These "bottom-up" budgets form the basis for the overall monitoring and evaluation effort. In addition, each program's share of the total budget, total projected participation and total projected avoided costs for all programs were considered. These considerations helped assure that the proposed approach balances various requirements.

The originally filed Iowa-Illinois plan included a total budget of $17,309,082. Of this amount, $15,920,874 was for residential and non-residential programs. Of that amount, $2,785,000 (17.5 percent) was budgeted for monitoring and evaluation. Iowa-Illinois has come to believe that excellent monitoring and evaluation results may be provided with a significantly reduced budget. In this plan the total budget for monitoring and evaluation is set at $1,172,638. This amount represents 7.37 percent of the total budget for residential and non-residential program implementation. In addition, $224,500 is budgeted for acquiring meters. This amount is 1.41 percent of the total for program implementation. The balance of the budget originally designated for monitoring and evaluation will be used for programs.

The rest of this section reviews the primary considerations; the perspectives gained by reviewing each program's share of total budgets, participation, and avoided costs; and presents budgets for each year.

3.2 Primary Requirements

In developing comprehensive monitoring and evaluation plans, it is necessary to balance efforts across three types of evaluation, to balance efforts over program life cycles, and to assure that minimum activity levels can produce useful results. The five primary requirements and their impacts on budgets include the following:

1. The key concern for Iowa-Illinois, their ratepayers, stockholders, and the citizens of Iowa is determining the cost-effectiveness of the Iowa-Illinois programs. This requires careful planning and execution of impact evaluations of each program with specific analysis of each measure.

Budgets must be provided to obtain sound estimates energy savings and demand reductions.

2. A second concern is assessing the persistence of the savings identified in the period immediately after implementation. Thus, the IUB requires the monitoring and evaluation cover the three years of implementation of the current energy efficiency plan, plus two post-plan years.

   Budgets must cover monitoring and evaluation activities starting in 1992 and continuing until the end of 1996.

3. The degree to which Iowa-Illinois' programs reach eligible customer groups and provide satisfactory services is also of concern.

   Budgets must also provide adequate funds for market assessments and process evaluations.

4. Program activities may be expected to follow a development cycle resembling a logistic growth curve. Thus, during the start-up period participation is modest, during a growth period of full acceptance participation reaches a peak, and, when most interested customers have been served, participation decreases.

   These planned budgets and evaluation activities reflect the start-up and growth phases of Iowa-Illinois' programs.

5. Finally, there are some minimum costs to conduct evaluation activities. Program managers and regulators always have a few questions to be answered for even the smallest programs. Just for a program with modest participation and impacts, there are minimum amounts that must be budgeted for reviewing program history, for reviewing the engineering estimates of savings, or for compiling and analyzing billing histories.

   These minimum levels may cause the monitoring and evaluation budget for small programs to appear larger than may be warranted in comparison to those programs' share of the total program budget.

3.3 Other Considerations in Allocating Budgets

Tables 3-1 through 3-3 display each program's share of the total budget, projected participation, and projected avoided costs for each program year and for the full three year plan. In addition to the percentage of the totals, a rank based on the three year values is shown for each program.

These tables show the programs from different perspectives. Each perspective contributes to allocating budgets across programs.

- The shares of the total program budgets indicate relative levels of utility activity in each program. Activity levels have been used to check the allocation of process evaluation budgets across programs.

- The rankings and shares related to participation have been used in apportioning market assessment and process evaluation budgets. However, due to the minimum budgets required to conduct data collection, the 90:10 split between residential and non-residential participation has been modified. All programs have been allocated sufficient funds to conduct appropriate process evaluations.

- The rankings and shares related to avoided costs have been used in apportioning impact evaluation budgets across programs. The avoided costs are split very evenly between the total for all residential programs and the total for all non-residential programs. Thus, the overall impact budgets have also been split quite evenly to provide sufficient funds to address the more complex impact analyses required in the non-residential programs.

Table 3-3
Program's Share of Total Avoided Costs, By Year

Program	1992	1993	1994	1992-1994	Rank
RESIDENTIAL					
Weatherization	31.9	29.3	27.1	29.2	1
H-E Equip. Rebates	14.6	13.9	12.7	13.8	3
Low Income	0.4	0.4	0.4	0.4	8
New Construction	5.0	5.95	6.20	5.7	6
Residential Total	**51.9**	**49.5**	**46.3**	**49.2**	
NON-RESIDENTIAL					
Shopping List Rebates	30.4	29.7	29.5	29.0	2
Customized Rebates	7.7	9.1	9.4	8.9	4
Ind. Process Optimization	7.0	8.3	9.7	8.8	5
New Bldg. Construction	0.0	0.3	1.1	0.7	7
Non-Residential Totals	48.1	50.5	53.7	50.8	

3.4 Monitoring and Evaluation Budgets

The detailed breakdown for each calendar year and for the total of all activities from 1992 through 1996 are presented in the following five tables.

The total budget of $1,172,638 for evaluating the 1992 to 1994 programs has been allocated in the following manner:

- 53 percent to residential programs and 47 percent to non-residential programs;
- 45 percent to data collection, 25 percent to impact evaluation, 11 percent to process evaluation, 10 percent to reporting, 5 percent to market assessment, and 5 percent to ongoing planning and evaluation management;
- the allocation over the five year period includes 27 percent for 1992, 32 percent for 1993, 31 percent for 1994, and 10 percent for 1995 and 1996.

In addition to the budgets detailed in these tables, $224,500 is budgeted for acquiring meters and related equipment.

Monitoring and Evaluation Budget for Iowa-Illinois Gas and Electric Company Energy-Efficiency Programs (Dollars)

SECTOR Program	1992 Planning and Management	Data Collection and Metering	Market Assessment	Process Evaluation	Impact Evaluation	Reporting	1992 TOTALS	Pct. of Total
RESIDENTIAL								
Weatherization	3,000	16,700	3,000	10,000	12,500	4,000	49,200	16%
H-E Equip. Rebates	3,000	21,800	3,000	10,000	12,500	4,000	54,300	17%
Low Income	500	2,475	500	2,500	1,250	2,000	9,225	3%
New Construction	1,200	14,500	2,000	7,000	3,750	2,000	30,450	10%
Master Non Part. Survey	1,000	25,000	0	0	0	4,000	30,000	
Residential Totals	8,700	80,475	8,500	29,500	30,000	16,000	173,175	55%
NON RESIDENTIAL								
Shopping List Rebates	3,000	23,000	3,000	6,000	12,500	4,000	51,500	16%
Customized Rebates	1,000	5,000	2,000	4,000	6,250	2,000	20,250	6%
Ind. Process Opt. Rebates	1,000	8,000	2,000	4,000	6,250	2,000	23,250	7%
New Bldg. Construction	1,000	9,000	2,000	4,000	6,250	2,000	24,250	8%
Master Non Part. Survey	1,000	20,000	0	0	0	3,000	24,000	8%
Non Residential Totals	7,000	65,000	9,000	18,000	31,250	13,000	143,250	45%
Overall Totals	$15,700	$145,475	$17,500	$47,500	$61,250	$29,000	$316,425	100%
	5%	46%	6%	15%	19%	9%	100%	

Monitoring and Evaluation Budget for Iowa-Illinois Gas and Electric Company Energy-Efficiency Programs (Dollars)

1994

SECTOR Program	Planning and Management	Data Collection and Metering	Market Assessment	Process Evaluation	Impact Evaluation	Reporting & Filing	1994 TOTALS	Pct. of Total
RESIDENTIAL								
Weatherization	3,308	16,000	3,308	10,000	16,500	6,000	55,115	15%
H-E Equip. Rebates	3,308	15,800	3,308	10,000	16,500	6,000	54,915	15%
Low Income	551	4,250	551	2,750	3,300	3,000	14,403	4%
New Construction	1,323	15,000	2,205	7,500	8,250	3,000	37,278	10%
Master Non Part. Survey	1,103	25,000	0	0	0	6,000	32,103	
Residential Totals	9,592	76,050	9,371	30,250	44,550	24,000	193,813	53%
NON RESIDENTIAL								
Shopping List Rebates	3,308	26,750	3,308	6,250	16,500	6,000	62,115	17%
Customized Rebates	1,103	5,000	2,205	3,750	8,250	3,000	23,308	6%
Ind. Process Opt. Rebates	1,103	5,000	2,205	3,750	16,500	3,000	31,558	9%
New Bldg. Construction	1,103	8,700	2,205	3,750	8,250	3,000	27,008	7%
Master Non Part. Survey	1,103	20,000	0	0	0	4,500	25,603	7%
Non Residential Totals	7,718	65,450	9,923	17,500	49,500	19,500	169,590	47%
Overall Totals	$17,309	$141,500	$19,294	$47,750	$94,050	$43,500	$363,403	100%
	5%	39%	5%	13%	26%	12%	100%	

Monitoring and Evaluation Budget for Iowa-Illinois Gas and Electric Company Energy-Efficiency Programs (Dollars)

SECTOR Program	1995 and 1996 Persistence of 1992 to 1994 Program Activities							
	Planning and Management	Data Collection and Metering	Market Assessment	Process Evaluation	Impact Evaluation	Reporting	1995 & 1996 TOTALS	Pct. of Total
RESIDENTIAL								
Weatherization	500	3,000	0	0	7,500	2,500	13,500	12%
H-E Equip. Rebates	500	3,000	0	0	7,500	2,500	13,500	12%
Low Income	500	3,000	0	0	2,500	1,500	7,500	6%
New Construction	500	4,000	0	0	5,000	2,000	11,500	10%
Master Non Part. Survey	0	0	0	0	0	0	0	
Residential Totals	2,000	13,000	0	0	22,500	8,500	46,000	40%
NON RESIDENTIAL								
Shopping List Rebates	500	12,600	0	0	7,500	2,500	23,100	20%
Customized Rebates	500	7,000	0	0	5,000	2,000	14,500	13%
Ind. Process Opt. Rebates	500	7,000	0	0	5,000	2,000	14,500	13%
New Bldg. Construction	500	7,000	0	0	7,500	2,500	17,500	15%
Master Non Part. Survey	0	0	0	0	0	0	0	0%
Non Residential Totals	2,000	33,600	0	0	25,000	9,000	69,600	60%
Overall Totals	$4,000	$46,600	$0	$0	$47,500	$17,500	$115,600	100%
	3%	40%	0%	0%	41%	15%	100%	

Monitoring and Evaluation Budget for Iowa-Illinois Gas and Electric Company Energy-Efficiency Programs (Dollars)

1992 to 1996 Totals for Evaluation of 1992 to 1994 Programs

SECTOR Program	Planning and Management	Data Collection and Metering	Market Assessment	Process Evaluation	Impact Evaluation	Reporting	1992 to 1996 TOTALS	Pct. of Total
RESIDENTIAL								
Weatherization	9,958	60,400	9,458	28,000	51,500	16,500	175,815	15%
H-E Equip. Rebates	9,958	70,400	9,458	28,000	51,500	16,500	185,815	16%
Low Income	2,076	20,325	1,576	7,250	10,050	8,500	49,778	4%
New Construction	4,283	56,500	6,305	20,500	24,500	9,000	121,088	10%
Master Non Part. Survey	3,153	75,000	0	0	0	14,000	92,153	
Residential Totals	29,427	282,625	26,796	83,750	137,550	64,500	624,648	53%
NON RESIDENTIAL								
Shopping List Rebates	9,958	97,300	9,458	17,250	51,500	16,500	201,965	17%
Customized Rebates	3,653	22,000	6,305	10,750	27,000	9,000	78,708	7%
Ind. Process Opt. Rebates	3,653	25,000	6,305	10,750	42,750	9,000	97,458	8%
New Bldg. Construction	3,653	36,500	6,305	10,750	29,500	9,500	96,208	8%
Master Non Part. Survey	3,153	60,000	0	0	0	10,500	73,653	6%
Non Residential Totals	24,068	240,800	28,373	49,500	150,750	54,500	547,990	47%
Overall Totals	$53,494	$523,425	$55,169	$133,250	$288,300	$119,000	$1,172,638	100%
	5%	45%	5%	11%	25%	10%	100%	

Total M&E Budget	$1,172,638	7.37%
Total Metering Capital	$224,500	1.41%
Total Program Budget	$15,920,874	

4.2 Evaluation Questions

4.2.1 Market Assessment

Baseline Research Questions and Issues

- Of the eligible housing (Iowa-Illinois provides the heating fuel), what proportion is:
  - single family homes?
  - mobile homes?
  - two- to four- family homes? and
  - multifamily buildings?
- What are the proportions of each housing type for those with less than R20 ceiling insulation?
- What was the annual incidence of installation of attic insulation up to R20 prior to the existence of the program? What factors influenced this market before Iowa-Illinois began its program?
- What are the baseline penetrations of the measures to be directly installed by Iowa-Illinois? What is the current state of the market for these measures?
- What percentage of the target population previously received an energy audit? (If possible, provide audit information by year).

4.2.2 Process Evaluation

Customer Response to the Program

- How aware and informed are Iowa-Illinois customers of this and other energy-efficiency programs?
- What are the relationships between customer participation in the audit, installation of the free measures, and customer participation in the clock thermostat and insulation options of the program? Which options of the program are most attractive to customers? What drives participation?

- Did customers find participation in each aspect of the program reasonably convenient? What rules or procedures posed barriers to increased participation? Was the direct subsidy (for insulation) amount and structure appropriate for the program and the customers?

- What was the participation rate for the program? Does the participation rate and initial and final R-values for the insulation subsidy portion satisfy the cost-effectiveness requirements of the program?

- What were the primary reasons for nonparticipation in the audit? What portion of the customers declined the direct installation portion of the program? Why? What were the reasons for nonparticipation in the insulation rebate? Are any market segments especially unlikely to participate (e.g. owners of multifamily buildings)? What changes could be made to optimize participation?

Use and Satisfaction With the Products

- For each directly installed measure, what is the distribution of the number of units installed per household? How does this compare to program estimates? Do staff (energy specialists/installers) report that different numbers of measures are needed?

- What percentage of the measures installed were removed or never used? How do these estimates compare with program estimates?

- Why did customers remove or never use some measures that had been installed?

- Are customers satisfied with the products? What particular features do they find favorable or unfavorable?

- Do customers want or need more than a single compact fluorescent? What does Iowa-Illinois intend to achieve with the installation of a single compact fluorescent? Is Iowa-Illinois' expectation being met? Will Iowa-Illinois introduce additional incentives to move customers toward using more of these bulbs?

- What does Iowa-Illinois intend to achieve with the installation of the clock thermostat? Is Iowa-Illinois' expectation being met?
- To what degree are the customers sensitive to the rebate levels for insulation? Are customers satisfied with the amount of the rebate?

Free Riders and Free Drivers

- What was the participant's level of knowledge and adoption of measures prior to contact with the program?
- Were customers planning to purchase the targeted measures over each program year even if no program existed?
- What was the level of measure knowledge and adoption among non-participants?

Program Administration

- Is management of the program well organized? Are the goals clearly stated? Are roles and procedures well defined and documented? Is there clear control and accountability for expenditures? How are program expenditures documented and authorized?
- Do management and program field staff have a shared understanding of program goals and objectives?
- Are communications with the program contractors, vendors and their allies operating smoothly?
- Are the contractor assessors/installers adequately trained? Are assessors trained in maximizing energy efficiency opportunities during in-home visits?
- Is there adequate quality control of the direct installations? Is there a mechanism to provide feedback on problems noted in quality control inspections to program training activities?

- Is the list of participating installation contractors kept up to date? What responsibilities and standards do participating contractors agree to abide by? How do Iowa-Illinois and the program contractor assure insulation contractor compliance?
- Do program tracking systems adequately support customer service, program management and all monitoring and evaluation requirements?
- What are the costs of the program by the following major categories: administration, advertising and promotion, monitoring and evaluation, installed equipment costs, and customer incentives?

4.2.3 *Impact Evaluation*

- What reductions were achieved in energy and demand?
- What are the free rider and free driver effects? What are the net effects of these factors on program reductions?
- To what extent would customers who chose to participate have been expected to save more or less than non-participants, even without the program?
- What are the net effects of changes in weather and economic conditions?
- What is the level of persistence of program measures and savings?
- What degree of takeback (snapback) is evidenced by program participants?

4.3 Technical Approach

4.3.1 Introduction

There are three aspects to the program evaluation for the Residential Weatherization Program:

- Market assessment;
- Process evaluation; and
- Impact evaluation.

These evaluation activities are supported by data collection activities, which are described in section 4.3.5.

4.3.2 Market Assessment

The market assessment will be performed in conjunction with other residential program market assessments. This will establish a baseline perspective of the current residential energy efficiency market, specific to the program measures offered in this program. The market assessment will be based on the information from the participant and nonparticipant surveys. This market assessment will provide data to estimate the net effect on the program impact of behavioral influences such as free-ridership, and self-selection bias. The baseline market assessment will also provide information on customer attitudes and responses from the market subsectors. It is likely, for example, that owners of multifamily buildings and mobile homes will have different perceptions of the need for specific energy-efficient technologies, based not only on the differences in the types of buildings they operate, but also on differing responsibilities for paying for energy costs. Finally, the market assessment will be a source of information on the saturation of energy-efficient technologies and measures in the population of interest.

4.3.3 Process Evaluation

The process evaluation will examine the program's operation and qualitatively measure its effectiveness. Program optimization recommendations will then be made for all levels of the program. The on-going evaluation process will enable Iowa-Illinois to get quick feedback during the critical early months of the program and to implement changes that will immediately affect program operations. The program activities described below are oriented toward delivering rapid feedback which will be of immediate use to program managers.

Implementation Analysis. The implementation analysis will accomplish two objectives for both the process and impact evaluations:

1. The program database will be analyzed to determine if it contains all the critical data needed for Iowa-Illinois to understand and manage the program.

2. The implementation analysis will provide estimates of parameters, determined in consultation with Iowa-Illinois that are contained in the database.

The combination of activities will provide an on-going understanding of program accomplishments. For example, savings estimates will be derived from the quantities of measures installed by contractors. Program participant profiles will be generated from demographic data, and a detailed census of measures installed by the program will be determined.

4.3.4 Impact Evaluation

The impact methodology for the Residential Weatherization program follows the general procedure described in Section 2.

Task 1. Select the Sample. The participant and nonparticipant samples are taken as components of the general sample described in Section 2. Participation in all programs, and dates of participation, are identified. The proportions of Weatherization program participants who also participated in other programs will be estimated. For those participating in other programs, the distribution of a length of the time between the participation dates will also be examined.

Task 2. Collect Survey Data. Most of the questionnaire administered to participants and nonparticipants will be the same as that administered to the general sample. Some additional questions will focus on weatherization measures undertaken, both within and outside of the program, and major equipment replaced. The dates of these actions will also be collected.

Task 3. Conduct On-site Audits. This task will be performed as described in Section 2.

Task 4. Refine the Engineering Models. This task will be performed as described in Section 2.

Task 5. Estimate Savings Using Engineering Models and Conduct Implementation Analysis. This task will be performed as described in Section 2.

Task 6. Conduct Billing Analysis (Stage 1). This task will be performed as described in Section 2.

Task 7. Conduct Multivariate Regression Analysis (Stage 2). This task will be conducted according to the steps described in Section 2. For the participation modeling (Step 1), participation in all applicable programs will be modeled jointly. In addition, the participation in the weatherization program will be modeled as a nested decision. That is, the choice is made first to participate in the direct-install and assessment component. The choice to participate in the subsidized insulation component can only be made if the first component is chosen.

Task 8. Conduct Demand Savings Analysis. The demand analysis will be conducted according to the procedures outlined in Section 2. For the Weatherization program, the load data required will be obtained from the utility's current load research data.

4.3.5 Data Collection

A coordinated set of data collection activities will be undertaken to inform the various analysis activities. These activities include:

- records and tracking system reviews;
- in-depth interviews with program staff and contractors;
- participant and non-participant mail or telephone surveys;

- billing history compilation; and
- metering (as appropriate).

Both quantitative and qualitative data will be collected. Each component of the evaluation will draw on more than one data gathering activity. For example, some behavioral net effects on gross program impacts will be determined from the results of customer survey questions concerning attitudes toward energy conservation, free riders and free drivers, and whether measures installed in participants' homes are retained and used.

Baseline Information. A stratified sample for a telephone survey of Iowa-Illinois customers will provide baseline data for the program evaluation. The sample selection and questionnaire design will be undertaken as soon as possible. The final sample size will be 1,200 completions. The data gathered in these surveys will be supplemented by on-going participant follow-up surveys.

Participant Satisfaction Survey. A second data collection effort will be a mail survey of all program participants. This survey will principally examine customer satisfaction with the program measures, the insulation subsidy strategy and the program process. This type of survey is inexpensive to conduct and has an expectation of high rates of response, based upon similar work done for other utilities in Iowa and in other states.

This survey will be performed within 30 to 45 days after the energy audit (while participants' memories are still fresh). The survey will include customer satisfaction questions (Was the audit conducted in a professional manner? Were you satisfied with the operation of the low-flow showerhead?) and will also question the customer concerning intentions to participate in the insulation subsidy. In addition, this survey will be designed to be sent to those customers who take advantage of the direct subsidy for insulation rebate. This satisfaction tracking will provide a quick means for rapid feedback about customer satisfaction and give Iowa-Illinois the opportunity to make on-going adjustments for problems identified in the responses. These surveys will also address free-ridership issues and customer costs while this information is easy for customers to recall.

Participant and Nonparticipant Surveys. At the end of the Year 1, there will be a participant/nonparticipant telephone survey, selected on a representative basis. This survey will collect information about customer attitudes and behavior. This information will be used to inform the process evaluation and to provide data for estimating behavioral parameters of the impact evaluation, such as persistence and free-ridership. Typical questions will target energy end uses; appliance types; usage patterns; customer knowledge of, and attitudes about, energy efficiency; actual and planned energy efficiency actions; and demographic characteristics.

Process Evaluation Surveys and Interviews. Data collected for process evaluation purposes will consist of customer satisfaction data garnered from participant surveys described above and interviews of Iowa-Illinois program staff, contractor managers and staff, and insulation contractor trade allies. Iowa-Illinois' input and approval will be obtained in the development of all interview protocols.

Four Iowa-Illinois central office and district staff will be interviewed shortly after program initiation and again at a later time. These interviews will focus on program development, information systems, reporting, quality control and assurance and other contractor issues. Interview schedules will be adjusted to account for staggered program startups, as needed.

The turnkey contractor will be interviewed concerning the operation of the program. Issues such as contractor organization, job scheduling, information systems, auditor training, quality control, inventory procedures, problems encountered and problem resolution will be addressed. The contractor manager will also be asked about issues relating to Iowa-Illinois, such as clarity of performance standards, communication, training, and changes in program procedures. Contractor assessors will be asked questions about training, scheduling, typical installations, problems encountered on the job, and so on.

Insulation contractors participating in the rebate portion will also be interviewed. Two or three contractors will be interviewed in the first six months of the program. Two or three more participating contractors will be interviewed at the 10 month mark. Contractors' interview topics will include their perceptions of the effectiveness of the program, the attractiveness of the program to customers, ease of program operation, adequacy of rebates, timeliness, and possible changes in program design or implementation. Two to three contractors who have chosen not to participate in Iowa-Illinois' program will be interviewed concerning their reasons for non-participation.

Customer satisfaction data will be collected in the on-going surveys, annual participant surveys, interviews with program staff and contractors and in reviews of program records regarding complaint recording and resolution. Data to be gathered include customer acceptance of, and satisfaction with, the products offered, the installation of directly installed measures, quantities and types of complaints about program contractors and complaint resolutions logged by Iowa-Illinois.

Impact Evaluation and Metering Plans. Impact evaluation data collection will be undertaken for the purpose of determining net and gross program impacts. The primary data source for the Stage-1 impact analysis described in Section 2 will be customers' bills. These will be linked to weather data specific to the Iowa-Illinois' service territory. The weather data will be compiled from the National Oceanic and Atmospheric Administration's (NOAA) records.

The Stage-2 analysis will relate the weather-adjusted estimates of gross savings to information from the participant and non-participant surveys. These surveys will be designed to meet the needs of both the impact and the process evaluation analyses. Additional factors that may be included in the analysis are external factors such as level of economic activity in the service area in the pre- or post-program periods.

If appropriate, Iowa-Illinois will develop a metering plan and collect metering data on a sample of homes treated in the program. This approach does not immediately appear cost-effective because of the cost of acquiring a sufficient base of data to generalize the results to program participants, or even less likely, to the population eligible for the program. Iowa-Illinois will continue to review the benefits and costs of obtaining meter data for this program as the program progresses and other forms of impact analysis are undertaken.

5.1 Program Description

The Residential High-Efficiency Equipment Rebate Program offers rebates to residential customers who purchase the following types of high-efficiency equipment:

- Central Air Conditioners;
- Room Air Conditioners;
- Heat Pumps (Air Source, Ground-Coupled, Ground Water);
- Gas Furnaces or Boilers;
- Electric Water Heaters;
- Gas Water Heaters; and
- Programmable Clock Thermostats (Cool only, Heat and Cool, Heat Only)

Eligible customers include all existing residential customers who buy electricity or gas from Iowa-Illinois. Rebates are provided for end uses for which the fuel is provided by Iowa-Illinois. The equipment will be marketed through a trade ally network of retail vendors and heating ventilation and air conditioning (HVAC) contractors. Rebate applications are provided to customers by the trade allies. Rebates are intended to cover approximately two-thirds of the incremental cost between standard-efficiency and high-efficiency equipment in each measure category.

Program goals for years 1 through 3 are indicated in Table 5-1.

Table 5-1
Residential High Efficiency Equipment Rebate
Year 1 - Year 3 Program Goals*

	1992		1993		1994	
	Gas	Electric	Gas	Electric	Gas	Electric
Participation Goal	2,504	676	4,176	1,683	5,526	2,224
Energy Savings	184,067 Th	285 MWh	491,042 Th	994 MWh	897,317 Th	1,931 MWh
Demand Savings	1,568 Th/day Peak Day	148 kW Summer On-Peak	4,184 Th Peak Day	515 kW Summer On-Peak	7,645 Th Peak Day	1,000 kW Summer On-Peak

** Note: Participation figures show new participants for each year. Energy and demand savings figures show the savings in each year for the cumulative number of participants.*

5.2 Evaluation Questions

The questions listed below are particularly relevant to the Residential High-Efficiency Equipment Rebate Program. Additional issues which are applicable to all Iowa-Illinois energy efficiency programs are contained in Section 2.1, and will be incorporated in the evaluation of this program.

5.2.1 Market Assessment

- What is the current state of the market? What are the baseline saturations of standard-efficiency and high-efficiency measures offered by the Iowa-Illinois program among the eligible customer population?
- What are the baseline efficiencies for each type of equipment?
- What market segments are discernible?
- How do the markets this program addresses differ from those addressed by the new construction program?
- Were rebate levels set to reflect approximately two-thirds of the incremental cost between standard- and high- efficiency equipment? If the rebates were more or less than two-thirds of the incremental cost, did it affect participation rates? Did the rebates affect market prices for standard- and high-efficiency equipment? How?
- When are ordering and stocking decisions made by trade allies? How might Iowa-Illinois' decisions to change measures being rebated affect trade allies? What is the best time of year to announce such changes?
- What is the disposition of equipment replaced under this program? (See Section 2 for a discussion of replaced equipment.)
- Are there other types of equipment that should be included in this program? What are they?

5.2.2 Process Evaluation

- What were the primary issues considered in planning the program?
- What problems or concerns emerged during the program start-up period? How were they addressed? What additional operational issues have emerged?
- Are customers and trade allies aware of the program? How do most learn of it?
- Is the program reaching targeted customers at the rates anticipated? Has Iowa-Illinois changed its marketing of the program as a result of the response?
- Do participating customers differ from nonparticipants? How?
- Are the application forms readily available to customers? Are the forms clear and easy to complete?
- Has Iowa-Illinois provided rebates in the time promised? Is that interval adequate? How might the utility's rebate processing time be improved?
- What were customers' and trade allies' expectations of the program? Were they met?
- Do program requirements support or inhibit participation?
- What are the levels of free riders, free drivers and snapback? What explains the incidence of these factors? How might Iowa-Illinois affect them?
- Is the current program design the most effective means to influence customers to purchase high efficiency equipment? What changes might be made to the program to increase its effectiveness?

5.2.3 Impact Evaluation

Basic impact evaluation questions are discussed in Section 2.1.3. The Residential High-Efficiency Equipment Rebate Program focuses on major household appliances. Persistence issues will be similar to those for the weatherization program, focused on high-efficiency equipment rebated under this program.

5.3 Technical Approach

5.3.1 Introduction

There are three aspects to the program evaluation of the Residential High-Efficiency Equipment Rebate:

- Market assessment;
- Process evaluation; and
- Impact evaluation.

The basic evaluation issues and methods are described in detail in Section 2. Evaluation activities are supported by the data collection activities described in Section 5.3.5.

5.3.2 Market Assessment

The market assessment will be performed in conjunction with other Iowa-Illinois residential program market assessments. This will establish a baseline perspective on the residential energy efficiency equipment market, specific to the measures offered for rebate in this program. The market assessment will be made from information gained from participant and non-participant surveys of customers. The market assessment will provide data to help estimate the net effects on the program impact of behavioral influences such as free-ridership, and self-selection bias. The baseline assessments will provide information on customer expectations and responses from market subsectors. There may be differences in the market expectations between customers who have central air conditioning and those who use room air conditioners, for example, when considering upgrading existing equipment or the purchase of new equipment.

5.3.3 Process Evaluation

The process evaluation will examine the program's operation and qualitatively measure its effectiveness. Program optimization recommendations will then be made for all levels of the program, as needed. An early program review will provide Iowa-Illinois with quick feedback on program start-up issues during the critical early months of the program and provide the opportunity to make program adjustments. Subsequent periodic examinations will provide Iowa-Illinois with measures of overall program progress, customer satisfaction, and with an understanding of on-going program operation issues.

Implementation Analysis. The implementation analysis will accomplish two objectives for the process evaluation:

1. The program database will be analyzed to determine if it contains all the critical data required for Iowa-Illinois to understand and manage the program.

2. The implementation analysis will provide estimates of key parameters, determined in consultation with Iowa-Illinois, that are contained in the database for the purpose of determining program progress.

The implementation analysis will contribute to the process evaluation by providing profiles of program participants, the frequency and amounts of rebates by measure, and initial estimates of program savings determined from the measures rebated.

The implementation analysis is also the basis of engineering estimates of energy savings and demand reductions made before billing analysis is possible.

5.3.4 Impact Evaluation

The impact methodology for the Residential High Efficiency Equipment program follows the general procedure described in Section 2.2.4, Impact Evaluation Tasks.

Task 1: Select the Sample. The participant and non-participant samples are taken as components of the general sample described in Section 2.3.2. Participation in all programs, and dates of participation are identified. The proportions of rebate program participants who also participated in another program will be estimated . For those participating in other programs, the distribution of the length of time between participation dates will also be examined.

The rebate program is designed to increase the efficiency of the equipment purchased for "natural replacement." That is, the program is not intended to encourage accelerated equipment replacement. The comparison group of interest is a set of customers who replaced equipment, but did not participate in the program. Only a small fraction of the master sample are likely to have replaced equipment during the program year. As a result, the comparison group may be very small, especially for some types of equipment.

One strategy to improve the size of the comparison group is to identify customers who made equipment purchases through trade allies. Customers so identified would be used as a supplementary sample, in conjunction with eligible customers from the master sample. Appropriate statistical weighting procedures would be required to ensure unbiased estimates from this combination sample.

Task 2: Collect Survey Data. Most of the questionnaire administered to participants and nonparticipants will be the same as that administered to the general sample. Some additional questions will focus on major equipment replaced, both within and outside the program. The dates of these actions will also be collected.

Task 3: Conduct On-site Audits. This task will be performed as described in Section 2.

Task 4: Refine the Engineering Model. This task will be performed as described in Section 2.

Task 5: Estimate Savings Using Engineering Models and Conduct Implementation Analysis. This task will be performed as described in Section 2,

Task 6: Conduct Billing Analysis (Stage 1). This task will be performed as described in Section 2.

Task 7: Conduct Multivariate Regression Analysis (Stage 2). This task will be performed as described in Section 2. For the participation modeling (Step 1) participation in all applicable programs will be modeled jointly.

Task 8: Conduct Demand Savings Analysis. The demand analysis will be conducted according to the procedures described in Section 2. For the rebate program, the load data required will be obtained from the utility's current load research data.

5.3.5 Data Collection

A coordinated set of data collection activities will be undertaken to provide data for the analysis activities. These activities include:

- Records and tracking system reviews;
- In-depth interviews with program staff and trade allies;
- Participant and non-participant telephone interviews;
- Billing history compilation; and
- Metering (as appropriate).

Both quantitative and qualitative data will be collected. Most components of the evaluation will draw upon more than one data source. For example, self-selection effects on program impacts will be estimated from information obtained from both the participant and non-participant interviews.

Baseline Information. A stratified sample of Iowa-Illinois customers will provide baseline data for the program evaluation. This survey will be conducted for all residential programs and will determine basic demographic data, such as building types, sizes, household composition, fuel types and equipment saturations. The final sample size will be 1,200 completed interviews in 1992.

Customer and Trade Ally Focus Groups. Participant and nonparticipant focus groups will explore customer attitudes about the rebate program format, the program design, program procedures and the level of rebates offered by Iowa-Illinois. These groups will provide qualitative data that will be used to define important issues and focus the quantitative research on these issues. Typical areas involve knowledge and understanding of the program, reasons for participation and nonparticipation, satisfaction with the rebate design and process.

An additional focus group of trade allies will examine program issues from the dealer and contractor perspectives. Typical areas are reasons for participation in Iowa-Illinois vendor program, understanding of how trade allies promote the program, the effects of the program on trade allies' inventory and stocking practices, alterations caused by the program in typical equipment sold and installed. Trade ally focus groups will be coordinated with other residential program research.

Telephone Survey: Participants and Nonparticipants. At the end of the first program year a survey will be conducted for a sample of participants and nonparticipants. This survey will seek basic demographic data, for comparison between the two groups. Survey question areas will include customer recognition of, and satisfaction with, the High-Efficiency Equipment Rebate program, reasons for participation and nonparticipation, free-ridership, free drivers, snapback and appropriateness of the measures offered for rebate. Additional areas will include customer knowledge of, and attitudes about, energy efficiency and customer actions that have been planned or taken.

In-depth Interviews. In-depth interviews will be conducted with key Iowa-Illinois program staff and with participating vendor and installer trade allies. The staff interviews will focus on program development, planning and implementation, the roles of staff and vendors, problems and opportunities for improvements, quality control and related program issues. Staff interviews will include four to six key Iowa-Illinois personnel.

On-site Verification Visits. Iow -Illinois will conduct 100 on-site verification visits. Qualified personnel will examine nameplate data for correlation with the information provided in the rebate application, and will also examine the installation and operation of the efficient equipment to determine that it is operating within the expected parameters. Visits will be scheduled to include representation of all measures offered in the program, including customer-installed weatherization measures.

6.1 Program Description

The Residential Low-Income Program is operated in conjunction with the state and federally-funded Weatherization Assistance Program (WAP). Operation of the program is performed by Community Action Agencies (CAAs) on a turnkey basis. All program activities are managed and conducted by the CAAs. The CAAs operating in Iowa-Illinois' service territory will provide energy efficiency services under the Weatherization Assistance Program and will also provide services to customers who are eligible for Iowa-Illinois' Low-Income Program. The Low Income Program provides the following:

- One compact fluorescent lamp;
- Hot water kit consisting of:

  - Water heater tank wrap;
  - Pipe insulation;
  - One low flow showerhead;
  - Two faucet aerators; and
  - Temperature setback to 120°F

These measures are provided and installed at no cost to the customer.

In addition to these measures, Iowa-Illinois will pay for two-thirds of the installed cost of attic insulation and band joist insulation for customers who buy their gas or electric heating fuel from Iowa-Illinois.

Customers must meet the WAP income eligibility standards and receive gas and/or electric from Iowa-Illinois to qualify for the program. Client selection for program participation is the same as the current client selection system designed by the U.S. Department of Energy. The CAAs give customers who have the highest heating bills the highest priority for service. Iowa-Illinois estimates that three percent of all residential customers are eligible for the program. The participation goals are found in Table 6-1.

Table 6-1
Residential Low-Income
Year 1 - Year 3 Program Goals*

	1992		1993		1994	
	Gas	Electric	Gas	Electric	Gas	Electric
Participation Goal	235	28	474	57	474	57
Energy Savings	25,768 Th	38.9 MWh	77,743 Th	117.8 MWh	129,717 Th	196.8 MWh
Demand Savings	126Th/day Winter Peak Day	9kW Summer On-Peak	380 Th/day Winter Peak-Day	28 kW Summer On-Peak	633 Th/day Winter Peak Day	47 kW Summer On-Peak

* *Note: Participation figures show new participants for each year. Energy and demand savings figures show the savings in each year for the cumulative number of participants.*

6.2 Evaluation Questions

6.2.1 Market Assessment

The Low-Income Program is being contracted as a turn-key program to local Community Action Agencies. Iowa-Illinois sends customers an annual bill insert describing the availability of energy services for low income people through the CAAs. Customers receiving program services will not be aware that Iowa-Illinois is partially funding the services they receive from the CAA. In consequence, marketing questions will be limited in scope to the following:

- What is the need for each of the covered measures?
- What is the annual capability to install the measures?
- Should additional measures be included?

6.2.2 Process Evaluation

The process evaluation will include the issues and methods described in Section 2 of this evaluation plan that are applicable to the Low Income Program. Because this program is administered on a turnkey basis, and because low income customers will not directly interact with Iowa-Illinois, the process evaluation will focus on the CAA contractors. Particular emphasis will be placed on the following areas:

- Do CAAs understand the Iowa-Illinois program (rules, restrictions, procedures)?
- How do CAAs deal with landlords ?
- Do CAAs have suggestions for improving the Iowa-Illinois program?
- What are the biases introduced by Iowa-Illinois' policy of not publicizing its role in the weatherization program?

6.2.3 Impact Evaluation

Impact evaluation is not within the scope of this evaluation plan because Iowa-Illinois is participating in the Statewide Low Income Collaborative Evaluation (SLICE) Project. Energy and demand reductions resulting from the Residential Low Income Program will be determined under the SLICE Project.

That project will also provide a basis for adjusting the statewide impacts to individual utilities. The Iowa-Illinois evaluator will use the results of the SLICE project to estimate the impacts for Iowa-Illinois' Low Income Project.

The evaluator will use these estimated impacts in cost-effectiveness calculations, similar to those that will be performed for the other Iowa-Illinois programs.

6.3 Technical Approach

There are two aspects to the evaluation of the Residential Low Income Program. Those aspects are:

- Market assessment; and
- Process evaluation.

The basic evaluation issues and methods are described in Section 2. Evaluation activities are supported by the data collection activities described in Section 6.3.4.

6.3.1 Market Assessment

The market assessment will be performed in common with other Iowa-Illinois residential program market assessments. This will establish a baseline perspective on the residential energy efficiency market, specific to the low income segment. The market assessment will be made from information gained from participant and nonparticipant surveys of customers and from CAA estimations of the low income market. Because the Residential Low Income Program will not be marketed in the typical manner, (CAAs will select the eligible low income customers with histories of high energy use) the market assessment is confined to estimating the need for energy-efficiency services in the low-income sector and

For this program, mostly qualitative data will be collected. Most components of the evaluation will draw upon more than one data source. A complete discussion of process evaluation issues that will be examined for all Iowa-Illinois programs is contained in Section 2.

Baseline Information. A stratified sample of Iowa-Illinois customers will provide baseline data for the program evaluation. This survey will be conducted for all residential programs and will determine basic demographic data, such as building types, sizes, household income level, household composition, fuel types and equipment saturations. The base sample size will be 1,200 completed interviews in 1992. An additional sample will be drawn from the low income population. Questions posed to low income customers will assess their awareness of weatherization programs and the factors that influenced them to apply for this type of assistance.

In-depth Interviews. In-depth interviews will be conducted with key Iowa-Illinois program staff and with participating CAA staff. Both sets of staff interviews will focus on program development, planning and implementation, the roles of staff and vendors, problems and opportunities for improvements, quality control and related program issues. Staff interviews will include four to six key Iowa-Illinois personnel. Contractor interviews will include all key participating CAA personnel.

On-site Verification Visits. Iowa-Illinois will conduct on-site verification visits with at least ten percent of the participants. Qualified personnel will examine the measures installed against the CAA program installation records for completeness and quality of installation. These visits are intended as a supplement to the quality control performed by the program contractor.

7.1 Program Description

The Iowa-Illinois Residential New Construction Program offers financial incentives to builders, developers, homeowners and contractors to install high efficiency measures into new residential dwellings. The program also incorporates an educational component on energy efficient construction directed to home builders.

At present the program offerings are similar to those of the residential high efficiency equipment rebate program. Under this program, both high efficiency gas and electric measures are covered. These measures include:

- High efficiency central air conditioners;
- High efficiency room air conditioners;
- High efficiency air and ground water source heat pumps;
- High efficiency water heaters;
- High efficiency furnaces and boilers; and
- Clock thermostats.

For each of these measures the minimum eligibility criteria are set above the federal appliance standard levels. The incentive payment is on a per unit basis for water heaters and is based on the improvement in efficiency for the HVAC measures. As a result, increasing incentive payments are paid for higher efficiency heating and cooling measures. The incentive for each piece of equipment is set at the same dollar amount as for a replacement under the High-Efficiency Equipment Rebate Program. Builders can usually buy equipment wholesale. Thus the rebate dollar amount designed to cover two-thirds of the incremental (retail) cost for existing buildings should cover nearly all of the incremental (wholesale) cost for new construction.

The current Residential New Construction Program design does not include incentives for thermal envelope measures (increased insulation, high performance glazing and air infiltration measures). It is anticipated that these measures will be incorporated into the program in late 1992. The minimum eligibility criteria for these measures may be based, in part, on the requirements of the proposed Iowa Home Energy Rating System (HERS).

The program will be promoted through a number of routes. Targeted groups include builders, developers, retail vendors, installers and homeowners. In addition, the final program design may incorporate an award component to increase awareness of energy efficient construction practices.

Program participation rates, expressed as a percentage of new construction starts, are expected to be 10 percent, 25 percent and 40 percent for 1992, 1993 and 1994 and beyond, respectively. Estimated participant numbers for the program's first three years are as shown in Table 7-1.

7.2 Evaluation Questions

7.2.1 Process Evaluation

In many respects, utility new construction program evaluations entail a more complex set of issues than residential retrofit programs. New construction program evaluations are complicated by a variety of factors, including:

- Multiple potential decision makers regarding program entry;
- Multiple marketing target audiences;
- Multiple measures installed per participant;
- Long lead times from program application to measure installation verification;
- Smaller participant and non-participant populations;
- No pre-participation (*before*) consumption baseline data.

While many of these concerns must also be addressed in retrofit program evaluations, they tend to be more significant in new construction evaluations.

Table 7-1
Residential New Construction
Year 1 - Year 3 Program Goals*

	1992		1993		1994	
	Gas	Electric	Gas	Electric	Gas	Electric
Participant Goal	149	24	373	60	596	96
Energy Savings	67,329 Th	30 MWh	235,633 Th	106 MWh	504,888 Th	226 MWh
Demand Savings	699 Th/day Winter Peak Day	12 kW Summer On-peak	2,447 Th/day Winter Peak Day	43 kW Summer On-Peak	5,242 Th/day Winter Peak Day	92 kW Summer On-Peak

* *Note: Participation figuires show new participants for each year, Energy and D emand savings figures show the savings in each year for the cumulative number of participants.*

The Iowa-Illinois Residential New Construction Program evaluation contractor must incorporate these concerns in formulating the evaluation's data collection requirements and in developing the primary program issues to be addressed by the evaluation. Many of the issues described in Section 2 will be addressed in the evaluation. Issues specific to Iowa-Illinois' Residential New Construction Program are listed below.

- **Program Development**

  What parties, e.g., utility, home builders, realtors, etc., were involved in creating the program? What effect did their input have on the current program design?

  How were program technical requirements and incentive levels developed?

- **Program Marketing and Delivery to Home Builders and Developers**

  What is the current marketing approach to reach home builders? Is it successful in generating interest and, ultimately, participation?

  Does the program meet home builders' needs and perceptions? What sells new homes?

  Do certain groups, e.g., custom home builders, need to be approached differently?

  Is home builder training adequate?

- **Program Marketing and Delivery to Realtors and Lenders**

  How are realtors and lenders being approached by Iowa-Illinois regarding this program?

  Do these parties see program participation as a means to distinguish qualifying homes?

  Are less stringent income/debt requirements being used for homeowners buying qualifying homes?

- **Program Marketing and Delivery to Home Buyers**

What sells homes to the buyers of new homes? How does this compare to program marketing efforts? Does the program need to be repositioned, e.g., healthy homes or quality construction vs. energy efficiency?

Do home buyers understand the program's benefits?

Are home buyers comfortable with the technologies being promoted and installed.

- **Free-ridership and Snapback Effects**

What are nonparticipant practices regarding efficiency measures?

What would participating home builders and contractors have installed in the absence of the program? What are they installing in nonparticipating new homes?

Is participant operating behavior, i.e., temperature setpoints and air conditioning run hours, different from non-participants?

7.2.2 Impact Evaluation

The objective of the Iowa-Illinois Residential New Construction Program is to increase the energy efficiency of new homes. The focus of the impact evaluation is different for a new construction program. Unlike residential retrofit programs, where the baseline or pre-installation consumption can be readily observed, there is no prior consumption data for new construction. This aspect creates a challenge for evaluators.

This section describes some of the issues that must be addressed when measuring the savings impacts attributable to a new construction program and the methodologies that can be utilized to estimate program impacts.

- **Identifying the Appropriate Baseline**

  Since no prior consumption data exist for the market segment, the appropriate, albeit hypothetical, baseline efficiency level must be identified. The misspecification of the baseline efficiency will produce inaccurate savings estimates.

- **The "Spill Over" Effect**

  One objective of new construction programs, and DSM programs in general, is to move the market toward a more energy efficient baseline. The presence of a new construction program may affect the type of building supplies and materials stocked by regional dealers. Therefore a nonparticipating dealer builder may purchase more efficient products than in the absence of the program. Also, customers building new homes may require their homes to meet some of the efficiency standards offered under the program but may choose not to participate in the program. These customers are known as free drivers.

7.3 Technical Approach

7.3.1 *Process Evaluation*

The evaluation of Iowa-Illinois' Residential New Construction Program will rely on both primary and secondary data sources. In general the data collection and analysis activities are similar to those described in Section 2. Therefore, only the activities unique to the residential new construction are described below.

- Review of program database and tracking system, including examination of selected program application forms.
- Interviews with utility program planning and implementation staff.
- Interviews with parties involved with program design or significantly affected by the program.
- Participant surveys, including both the on-going "satisfaction" census of participants and annual surveys.

- Non-participant surveys.
- Trade ally focus groups.

Additional data collection activities may be required depending on the final program design. For example, focus groups with lenders may be appropriate if the program addresses the increased affordability of energy-efficient housing as an important program benefit.

To the extent possible, data collection activities for the New Construction Program will be coordinated with those of other programs. In particular, data collection requirements for the High-Efficiency Equipment Rebate Program may overlap with those of the New Construction Program.

Data collection activities for the New Construction Program will be phased in over a two-year period. This reflects different evaluation task requirements, gradual program ramp-up, the duration of construction of a new home, and that the final program design will not be completed until the third quarter of 1992. Note that several of the data collection tasks will target gas and electric measures separately and that the findings will also be reported on a fuel type basis where appropriate.

Task 1: Review Program Development and Current Program Progress. A clear understanding of the development of a utility DSM program is important in assessing the program's current Residential New Construction DSM Program, given the number of different parties whose interrelationships are vital to the program's functioning and ultimate success. In-person interviews with individuals and groups involved in, or consulted during, the development of a new construction program will be conducted. These may include:

- State and regional home builder associations.
- Distributors and manufacturers of manufactured homes.
- State and local code officials.
- State energy office (DNR) and Iowa Utility Board (IUB).

- Utility program planning and implementation staff.
- State and local realtor associations.
- Equipment manufacturers and distributors.
- State banking associations.
- Federal and secondary mortgage agencies - Veterans Administration, Fannie Mae, Freddie Mac, FHA, etc.

Task 2: Examine Program Targeting and Delivery. In this task the results from the participant and nonparticipant surveys, the program interviews from Task 1, trade ally focus groups, and the marketing material review in Task 5 will be synthesized to determine if the program has correctly identified the program's market targets and whether these efforts have been successful in generating program participation.

Review of the new-home buyer participant and nonparticipant surveys will allow an accurate demographic characterization of both populations, including:

- Type of home - custom, speculative, stick built, manufactured, etc.
- Geographic distribution.
- Urban/suburban/rural distribution.
- Median income levels and age of head of household.
- Home size and price.
- First-time home buyer or not.

For participants, these data will be cross-tabulated by program measures installed. Additional analysis of these data will be performed to determine how the decision to participate was made and what were the primary factors influencing the decision. From the non-participant survey results, the analysis will determine reasons for not participating.

To better understand the role of trade allies in promoting program participation, focus groups will be held with participating and non-participating HVAC contractors and home builders. Additional focus groups with lenders and/or realtors may be held, depending on the program's final design and marketing plan. The focus groups will address both program marketing and program delivery.

Task 3: Identify Barriers to Participation. This task will review the results of prior data collection efforts and synthesize and summarize barriers, both real and perceived, to program participation. This task will be conducted as described in Section 2.

Task 4: Determine Overall Program Satisfaction. While the ultimate customer for this program is the home buyer, it is imperative that program satisfaction also be ascertained from home builders and HVAC contractors at a minimum, and possibly for realtors and lenders. As a result, similar but different questionnaires will be administered to each group. Consistency among the survey instruments will be assured to allow direct comparison of results, where appropriate. In addition, each respondent group will be queried regarding the performance of all the other groups.

Task 5: Determine Effectiveness of Marketing Materials. This task will serve as a continuation of the marketing review undertaken in Task 2. The goal will be to qualitatively and quantitatively assess the effectiveness of the program's marketing materials. This determination will be based on the home buyer participant and nonparticipant surveys and the interviews and focus groups with the home builders and related trade allies. These groups will be asked whether the program's benefits, as defined by Iowa-Illinois, are clearly stated and effectively conveyed by the marketing materials. Where different materials and different marketing approaches, e.g., radio, bill stuffers, etc., have been employed, the relative effectiveness of each will be determined. The evaluation contractor will also review the program's marketing budget to determine which marketing approach appears to be most cost-effective.

Task 6: Review Utility Program Management and Administration and Recommend Changes to Improve Program Efficiency. The evaluation contractor will review each step of the utility's administration of the residential new construction program. This review will also entail a thorough review of the program tracking database and program application and verification forms. The contractor will assess communication channels between utility staff and outside parties, as well as within the utility, paying particular attention to mechanisms that allow the program staff to pinpoint nascent problems and apply corrective action. Potential administrative bottlenecks will be identified and recommended changes suggested. Program staffing levels will be reviewed in relationship to the program's goals through the year 2000.

Task 7: Assess Program Free-Ridership and Snapback Effects. As described in Section 2, the evaluation contractor will employ a variety of methodologies to measure program free-ridership and snapback effects. In general, two primary approaches will be used to measure these effects. The first will survey participants and non-participants on measure choices and occupant behaviors, while the second will compare pre- and post-program activities. The difficulties in evaluating program free-ridership and snapback are compounded in a new construction program in that a pre-program condition does not effectively exist. Further, in a new construction program, the definition of participant and non-participant may vary depending on the measures installed and the program design. With the exception of custom built homes, it is unlikely that the home owner will be the primary decision maker on the efficiency of the HVAC equipment and thermal envelope. As a result, these questions will be directed to the builders and HVAC contractors.

The primary parameters to be measured to assess program snapback will be changes in temperature setpoints and cooling equipment operating hours. Participating and non-participating homeowners will be asked about their current practices and, if applicable, their practices in their prior residences. The results from these questions must be carefully interpreted to reflect changes in household size, operating schedules, presence of setback thermostats and HVAC system zoning. Participant home owners will also be asked directly if their behavior would be different in a nonparticipating home.

7.3.2 Impact Evaluation

The impact methodology for the Residential New Construction Program follows the general procedure detailed in Section 2. The main distinction for the program is that there is no pre-program condition to use as a basis of comparison to evaluate the net program effect. Instead, the energy consumption of the participants is compared directly to that of a nonparticipant new construction group. As a result, greater reliance is placed on the comparison group, and also on the engineering estimates.

Task 1. Select the Sample. This new construction participant and nonparticipant samples are taken as components of the general sample described in Section 2.

Task 2. Collect Survey Data. Most of the questionnaire administered to participant and nonparticipants will be the same as that administered to the general sample. Some additional questions will focus on decisions made for the new house design or purchase.

Task 3. Conduct On-Site Audits. For this program, on-site audits will be conducted at nonparticipant homes as well as for participants. This additional data collection is necessary to ensure the accuracy of the engineering models, in particular the baseline assumptions for new construction. It is expected that 100 on-site audits will be conducted each year.

Data for new baseline construction practices have not previously been collected for homes in Iowa-Illinois' service territory. The utility will specify baseline efficiency levels to be assumed for the engineering analysis. The on-site audit will provide the field data to compare the new construction practices of the market to the standards set by this program. The nonparticipant comparison group will represent the baseline construction practices of the market in the Iowa-Illinois service territory.

Task 4. Refine the Engineering Models. This task uses the on-site audit data to ensure the accuracy of the inputs to the engineering models. Of particular importance is the accuracy of the baseline efficiency assumptions. This information is obtained from the nonparticipant audits.

Task 5. Estimate Savings Using Engineering Models and Conduct Implementation Analysis. This will be done according to the procedures outlined in Section 2.

Task 6. Conduct Billing Analysis (Stage 1). The Stage-1 billing analysis will determine weather-adjusted annual consumption for each participant and each nonparticipant for the first year following construction. No "savings" is computed for each customer from the billing analysis, since there is no "pre-program" condition that can be measured directly from billing records.

Task 7. Conduct Multivariate Regression Analysis (Stage 2). The models used for this task will be similar to those applied in the other residential programs, except that the energy model will model energy directly, not energy savings. Some predictor variables may be relevant for new construction that will not be considered for the other programs.

Task 8. Conduct Demand Savings Analysis. The demand analysis will be conducted according to the procedures outlined in Section 2. For the New Construction Program, the load data will be obtained by metering a subsample of participants and nonparticipants.

8.1 Program Description

This program falls under the Non-residential Building Systems Rebate umbrella and offers partial cash rebates to Iowa-Illinois' commercial and industrial customers who purchase and install high-efficiency HVAC and lighting equipment. The program covers a prescribed "shopping list" of both electric and gas energy-efficiency measures for cooling, lighting and space heating. The list of measures eligible for rebates typically includes high-efficiency cooling equipment (packaged single-zone and rooftop units), high-efficiency air-source heat pumps, programmable thermostats for gas furnaces, energy-efficient fluorescent lamps, T-8 lamps and ballasts, electronic ballasts, occupancy sensors, daylighting controls, HID interior lighting and exit light retrofits.

Rebate amounts are structured so that the incentives paid reflect dollars per kW of reduced demand. Incentive amounts are calculated by using a formula to assess the measure's two-year payback, 80 percent of the avoided costs and 50 percent of the incremental capital cost. Estimated incentives for the direct rebates portion of the program are $89 per kW for cooling, $157 per kW for lighting, $264 per kW for heat pumps and $63 per participant for gas heating. All non-residential Iowa-Illinois customers are potentially eligible for program participation.

This program is designed to complement the Iowa Department of Natural Resources' local government, school, and hospital energy programs.

Goals for this program are shown in Table 8-1.

8.2 Evaluation Questions

The following describes the evaluation questions specific to the Shopping List Rebate Program.

8.2.1 Market Assessment

- What is the current state of the market? What are the baseline saturations of standard-efficiency and high-efficiency measures offered by the Iowa-Illinois program among the eligible customer population?

Table 8-1
Non-residential Building System
"Shopping List" Rebate Program Goals*

Technology		1992	1993	1994
Cooling	Participation	22	44	85
	Energy Savings	468 MWh	1,405	3,215
	Demand Savings	174 kW Summer on-Peak	522 kW Summer on-Peak	1,195 kW Summer on-Peak
Lighting	Participation	154	296	404
	Energy Savings	2,929 MWh	8,560 MWh	16,245 MWh
	Demand Savings	725 kW Summer on-Peak	2,118 kW Summer on-Peak	4,020 kW Summer on-Peak
Heat Pumps	Participation	15	20	23
	Energy Savings	204 MWh	477 MWh	790 MWh
	Demand Savings	24 kW Summer on-Peak	57 kW Summer on-Peak	95 kW Summer on-Peak
Gas Heating	Participation	265	407	555
	Energy Savings	43,884 Th	111,283 Th	203,191 Th
	Demand Savings	409 Th/day Winter Peak Day	1,037 Winter Peak Day	1,893 Winter Peak Day

* *Note: Participation figures show new participants for each year. Energy and Demand savings figures show the savings in each year for the cumulative number of participants*

•What are the baseline efficiencies for each type of

- What market segments are discernible?
- How do the markets this program addresses differ from those addressed by the new construction program?
- Were rebate levels set at levels reflecting approximately two-thirds of the incremental cost between standard- and high-efficiency equipment? If the rebates were more or less than two-thirds of the incremental cost, did it affect participation rates? Did the rebates affect market prices for standard- and high-efficiency equipment? How?
- When are ordering and stocking decisions made by trade allies? How might Iowa-Illinois' decisions to change measures being rebated affect trade allies? What is the best time of year to announce such changes?
- What is the disposition of equipment replaced under this program? (See Section 2 for a discussion of replaced equipment).
- Are there other types of equipment that should be included in this program? What are they?

8.2.2 Process Evaluation

Program Design

- How do rebate amounts compare to the incremental costs associated with purchasing and installing high-efficiency equipment?
- How well does the Non-residential Building Systems: "Shopping List" Rebate program complement the Iowa Department of Natural Resources' programs for local government, schools and hospitals?
- Are the eligible technologies appropriate?

8.2.3 Impact Evaluation

- What are the energy and demand savings attributable to the program?
- What are the effects of free riders and free drivers on program savings (gross versus net savings)?
- What are the effects of weather and economic conditions on program savings?
- What level of persistence is present in the program? Does it vary by measure or end use?
- What degree of takeback (snapback) is indicated by program participants?

8.3 Technical Approach

8.3.1 Market Assessment

Participant Survey. A follow-up survey of participants will be done to obtain immediate reactions to their experience with Iowa-Illinois and the Shopping List Rebate Program.

Participant and Non-Participant Surveys. Mail surveys will be conducted with participants and telephone/mail surveys with non-participants to provide data on the current market saturation for each technology rebated under this program. A baseline survey of non-residential customers is being considered to provide basic, pre-program appliance and equipment saturation data. The surveys will focus on a mix of commercial and industrial end-users.

Trade Ally Survey. We will identify and interview a sample of each of the trade ally groups, including design professionals and members of the distribution network. The interviews will be designed to determine patterns for specifying and selling high-efficiency versus standard efficiency equipment and how the program has changed the sales mix of high-efficiency versus standard-efficiency equipment.

Staff Assessment. During their interviews, the corporate and field staff will be asked their views on the markets for each technology currently covered by the Shopping List Rebates and for nominations for additions to the list.

8.3.2 Process Evaluation

All participants will be asked to complete a short mail questionnaire within six weeks of the completion of participation in the program. As explained above, this survey will focus on participant satisfaction and reasons for participating.

In addition, a combination of mail and on-site surveys will be conducted with program participants to provide both qualitative and quantitative assessments of the program. Customers will be surveyed regarding attitudes toward and knowledge of the program, energy efficiency oriented purchase decisions and self-selection for the program.

A sample of approximately 200 participants and nonparticipants will be conducted to determine their awareness of the program, their reasons for participating or not participating, and overall reaction to the program.

The nonparticipant sample will serve as a comparison group for the program participants. The purpose of the nonparticipant group is to account for the changes in customer behavior that are attributable to factors other than program participation, i.e., economic changes. To ensure that the comparison group is similar to the participants, the sample will include only customers who installed the particular technology. For example, the comparison group for lighting participants are customers who installed energy-efficient lighting but did not participate in the Shopping List Program. The comparison group of customers will be selected by conducting a brief telephone survey to screen-out nonparticipants who have not installed the corresponding measures.

Focus groups and telephone surveys with design professionals, contractors and distribution channel members will be conducted to examine their reactions to program mechanisms, their attitudes toward high-efficiency equipment and their attitudes toward cooperation with Iowa-Illinois in this program. Trade allies will be asked to elaborate on how useful the program is as a tool to help promote and sell high-efficiency equipment and to discuss the positive aspects of the program as well as areas that need improvement. One-half of the surveys will be conducted with participating trade allies and the other half with non-participants.

In-person interviews will be conducted with all staff members concerning the administrative and service aspects of the program. A sample of field representatives who promote the program to customers and trade allies will also be interviewed in person. They will be asked to comment on how valuable they believe the program to be as a DSM tool, on how they perceive the customers and trade allies to receive the program , what is working well in the program, what improvements need to be made in the program and whether or not changes need to be made at the corporate staff level to improve support to field program efforts.

The questionnaires will be designed so that each respondent's (customer trade allies and staff) entire experience with the program from initial program communications through rebate receipt will be examined. Most data collection will occur in the fourth quarter 1992 and in the fourth quarter 1993.

8.3.3 Impact Evaluation

The objective of the impact analysis is to measure the energy and demand savings attribute to the Iowa-Illinois Non-residential Shopping List Rebate Program. Impact evaluation requires establishing a framework to efficiently collect critical measurement data and enable staff to routinely track the performance of the program. Therefore, it is critical to develop a thorough evaluation plan early in the implementation phase.

This section describes the proposed impact methodology for the Iowa-Illinois Shopping List Rebate Program. The plan incorporates the Iowa-Illinois tracking system to provide preliminary gross program savings estimates and then utilizes a combination of metering, engineering and statistical models to refine estimates of gross and net savings. The impact methodology for the program follows the general impact approach described in Section 2.

Task 1: Implementation Analysis. To monitor the activity and performance of the NR-1 Programs, an implementation analysis of the data collected in the tracking systems will be performed. The database will be reviewed to ensure the data are complete and that a consistent set of decision rules are used to generate programs information. This will identify areas for future enhancements in data collection and will validate the quality of the data. This task will also enable the evaluation contractor to determine if data can be incorporated into the impact analysis or if other (either primary or secondary) data are required.

On a routing basis, preferably weekly or monthly, the Iowa-Illinois staff should generate program summary statistics in formats that are consistent and understandable.

The implementation analysis will produce, at a minimum:

- The number of "participants" (some customers may participate more than once);
- The number of installed measures by product category (i.e. electronic ballasts, reflectors, etc.);
- The engineering-estimated energy and demand savings by product category and by fuel type; and
- Program and participant costs.

The database will also be reviewed for data integrity. The data collected in the database are critical for assuring that accurate , consistent, relevant and complete data are available for subsequent program evaluations.

The database will be reviewed for the following:

- Variables with a high proportion of missing values;
- outliers;
- inconsistencies in the magnitude of the data reported (i.e. annual operating hours of 10 hours) ; and

- ensure that customers with multiple meters have been identified and their billing data combined.

Task 2: Define wattage assumptions for installed and replaced lighting technologies. Wattage assumptions will be developed for the installed technologies offered under NR-1 and for the baseline technologies identified in the implementation plans. To ensure consistency, XENERGY recommends that wattage assumptions be obtained from a common data source such as the American National Standards Institute (ANSI).

Task 3: Estimate the Full Load Hours (FLH), Demand Factor (DF), Coincidence Factor (CF), and Interaction Factors (IF). FLH, DF, IF, and CF will be obtained from secondary data for the impact evaluation. Run meters will be installed to verify the demand factor assumptions and the participant survey will be used to determine the percentage of space conditioned floor space to calculate the interaction factors attributable to cooling bonuses and heating penalties..

Task 4: Estimate Engineering Algorithms. The engineering algorithms for lighting, air conditioning, heat pumps, and set back thermostats will be estimated to produce preliminary saving estimates.

Task 5: Conduct Billing Analysis (Stage 1). The billing histories for program participants and a sample of non-participants will be obtained from Iowa-Illinois' customer billing files. The data will be screened to remove anomalies from the data. The data will be weather normalized to control for weather sensitive loads. The data screening and weather normalization process will follow the approach described in Section 2.

Task 6: Conduct Multivariate Regression Analysis (Stage 2). Conditional demand and statistically adjusted engineering models will be specified and estimated to produce estimates of energy savings.

Conditional Demand Models. The explanatory variables will be based on date obtained from participant and non-participant surveys and may include:

- Implementing Run-time Meters; and
- On-site surveys.

Follow-up Survey. The follow-up census of participants will obtain immediate reactions to their experience with Iowa-Illinois and the Shopping List Rebates Program, information on their intentions to install energy-efficient measures without the program (free-ridership), reasons for participation, effectiveness of marketing efforts, expected savings, and customer costs.

Participant Non-Participant Surveys: Annual participant and non-participant surveys will be used to collect information related to program procedures, realization of savings, persistence take back and information on customer characteristics, attitudes and behaviors. These data are used for the process and impact evaluations.

Trade Ally Interviews: Interviews with trade allies will explore their reactions to the program, communications with Iowa-Illinois, their reported sales of energy-efficient equipment (if available), and their suggestions for program improvements.

Iowa-Illinois Staff Interviews: Interviews with Iowa-Illinois' staff will explore goal setting, program administration, their perception of trade ally and customer reactions, and their suggestions for program improvements.

Run-Time Meters: Meters that measure operating hours by period (i.e. peak and off-peak) will be installed in a sample of participants. Diversity, the percentage of fixtures on or off during a particular period, varies by technology and space type. Therefore a sampling plan will be specified to monitor the operating hours of specific installed technologies and space types (i.e. hallways, interior offices, store rooms, etc.). The results will produce diversity factors for each major building type. Run-time meters will be installed for 40 program participants.

On-Site Surveys: On-site surveys will be conducted for 40 program participants. The purpose of the visits will be to:

- Verify the measure installation and determine if the measure is performing as intended;

- To enhance the characterization of the participant population;
- To examine operating schedules and confirm impact algorithm assumptions;
- Assess the extent of market capture and identify lost opportunities; and
- Install run-time meters.

SECTION 9 NON-RESIDENTIAL CUSTOMIZED REBATE PROGRAM

9.1 Program Description

The Non-Residential Building Systems Customized Rebates Program is designed to encourage non-residential customers to purchase, install and operate energy-efficient building systems. Incentives are provided for customized packages of energy efficiency options that will largely be assembled by the customer. In some cases Iowa-Illinois may provide assistance in the form of partial funding for energy surveys and audits of customer facilities and the preparation of energy-efficiency improvement recommendations. The program potentially complements the Iowa Department of Natural Resources (DNR) programs for local governments, schools and hospitals.

The program includes both electric and gas energy-efficiency measures in the combinations service areas of Fort Dodge, Iowa City and the Iowa Quad Cities. Only gas energy-efficiency measures are included in the gas-only service areas of Cedar Rapids and Ottumwa. Any package of energy-efficiency measures that can demonstrate energy and capacity savings may be eligible for a cash incentive.

Typical measures that are commonly eligible include:

Electric Measures:
- High COP/multizone chiller systems
- Cooling towers
- Double-bundled chillers
- Ground-coupled heat pumps
- Efficient motors for fans
- Variable Air Volume (VAV) conversions
- Integrated lighting systems
- Efficient refrigeration and commercial appliances
- Advanced heat pump systems such as water loop heat pumps

Gas Measures:
- VAV conversions
- Insulated building shell
- Large high-efficiency boilers
- Large boiler conversions

The estimated incentive for the electric portion of the customized program is $54 per kilowatt. This is the average across a mix of measures and provides a simple payback for that mix of measures of

2.4 years. Incentives for both electric and gas measures are calculated as the smallest of three amounts: the amount required to buy the payback down to 2 years, 80 percent of the avoided cost, or 50 percent of the incremental capital cost. Projects with less than a 2 year payback receive no rebate.

The program's participation goals and estimated impacts are shown in Table 9-1. Over the three years included in this initial energy-efficiency plan, participation in the electric building system customized rebate program is projected at 43 participants and in the gas program at 44 participants.

Figure 9-1
Non-Residential Customized Rebate Program
Year 1-3 Goals and Estimated Impacts

PROGRAM		1992	1993	1994
Electric	Participation	6	15	22
	Energy Savings	1,045 MWh	3,656 MWh	7,486
	Demand Savings	315 kW Summer on-Peak	1,103 kW Summer on-Peak	2,258 kW Summer on-Peak
Gas	Participation	5	13	26
	Energy Savings	28,179 Th	101,444 Th	247,975 Th
	Demand Savings	282 Th/day Winter Peak Day	945 Th/day Winter Peak Day	2,310 Th/day Winter Peak Day

*Note: *Participation figures show new participants for each year. Energy and Demand savings figures show savings in each year for the cumulative number of peaticpants**

9.2 Evaluation Questions

9.2.1 Market Assessment

Due to the varying needs of individual large customers, it is difficult to assess the market for each of the many technologies and measures which may be covered by the customized rebates. The market assessment for some of the more popular items that are rebated through this program could involve explorations of the frequency, timing and extent of the building system replacements being made by typical customers. In addition, customers' plans for growth and building renovation and their other energy efficiency needs will be explored. Trade allies handling the popular items may also provide valuable insights into the extent of these markets so that Iowa-Illinois can develop a better sense of the appropriate level of effort for this program.

9.2.2 Process Evaluation

The process evaluation of the building systems customized rebate program will involve reviewing the experiences of a small number of participants (a projected maximum of 11 for 1992, 28 for 1993, and 48 for 1994). Each participant may have purchased different equipment or different combinations of equipment from one or more vendors. Some participants will also have had an energy assessment performed.

Thus, the key issues for this process evaluation will involve determining for these few participants:

- how clearly the Iowa-Illinois staff, the consulting engineers conducting assessments and the trade allies selling and servicing equipment understood the Iowa-Illinois program;
- how well utility staff, consulting engineers, and trade allies communicated among themselves and with customers;
- whether utility staff, consulting engineers, and trade allies provided services and products in a timely manner;
- whether application and approval procedures were easy to understand and complete;
- whether rebates were paid in a timely manner; and
- how well Iowa-Illinois staff, consulting engineers, and trade allies met customers' expectations.

9.2.3 Impact Evaluation

Due to the small numbers of participants projected for both the electric and gas building systems' customized rebates, little general information about energy savings and demand reductions can be developed during the period covered by the first energy-efficiency plan. Detailed information on the energy savings and demand reductions related to each specific installation will be developed by conducting individual case studies of each participant.

9.3 Technical Approach

9.3.1 Introduction

The basic approach for evaluating the small number of participants expected in the first year this program is offered will be to develop case studies for every participant. (A small number of participants may decline to be the subject of a full case study.) By the second and third years of the program, when 30 to 50 participants are projected, a non-random sample of approximately 10 participants will be selected. The sample would be based upon customer characteristics and measures installed. Case studies involve personalized data collection methods, including on-site surveys, in-person semi-structured interviews, and metering of specific equipment.

9.3.2 Market Assessment

To develop a better understanding of the natural replacement of building systems equipment, Iowa-Illinois will interview selected customers and trade allies about the market for equipment being rebated under this program. These interviews will provide qualitative information on the questions listed above. As warranted, some of these questions may be addressed by surveys of larger samples of customers. More general questions about customers' energy needs and plans for facility changes will be addressed.

9.3.3 Process Evaluation

Information about customer satisfaction with the process of participating in this Iowa-Illinois program will be obtained from follow-up interviews with each participant. These telephone or personal interviews will be semi-structured to allow customers the flexibility to mention issues beyond the follow-up interview topics being addressed for some of the larger programs.

Additional information about customer satisfaction with the equipment they have installed and other measures they have adopted will be gathered from the successive interviews that will be conducted with all participants who agree to be the subject of a case study.

9.3.4 *Impact Evaluation*

Customized Rebate program for commercial and industrial customers will be evaluated as a set of case studies. The sequence of tasks is still similar to those for the statistically based evaluations. Compared to the residential programs, the nonresidential programs have more emphasis on demand impacts, more emphasis on lighting measures, and less on temperature-sensitive loads. These differences affect the emphasis of the modeling, and of the data collection to support that modeling.

Because the actions taken under this program are tailored to the needs of a small number of diverse participants, the impacts will depend on the particular customers who participate. A detailed understanding of these customers will therefore be essential to a valid estimate of program impacts. The impact evaluation strategy for the tailored programs emphasizes on-site inspections. These inspections will provide accurate inputs for engineering calculations of savings. In addition, interval meters will be installed on major equipment predicted to be associated with large energy savings or demand reductions. The meters will be installed prior to the equipment installation to collect the before and after demand levels.

On-site visits will be conducted for all participants, or for a substantial fraction of participants, as participation increases in the second and third program years. If all participants cannot be visited, priority will be given to customers with larger anticipated savings. Likewise, the end-use metering will target the customers and the specific pieces of equipment associated with the highest predicted savings. Measures may also be targeted because there is greater uncertainty about their impacts and resolving the uncertainty is important to decisions about the future program.

Task 1. Select the Cases for Study: Up to 10 participants per program year will be selected for case studies. If more than 10 customers participate in a given year, a non-random sample will be selected, based on the distribution of characteristics among the participants. Factors such as SIC, annual consumption and peak demand, and utility district will be considered in selecting the sample for study.

For each participant, the dates when rebated equipment is to be installed will be determined. Some customers may participate in the program more than once, for different pieces of equipment. The multiple participation dates will be tracked.

Task 2. Collect Metering Data: Meters will be installed prior to the installation of rebated equipment. End-use meters will be installed for major pieces of equipment, and will operate both before and after new equipment is installed. Simple run-time meters will also be installed to collect hours of operation for lights.

Task 3. Conduct On-site Audits and Interviews: On-site audits and interviews will be conducted to determine detailed structural and operating characteristics of each case studied. Included in this data collection will be the key inputs required by the engineering models, as well as customer information required for the process evaluation.

Task 4. Refine the Engineering Models: This task will be performed as described in Section 2.

Task 5. Estimate Savings Using Engineering Models and Conduct Implementation Analysis: This task will be performed as described in Section 2.

Task 6. Conduct Billing Analysis (Stage 1): The billing analysis for these case studies will be exploratory in nature. Degree-day models similar to those used for the residential evaluations will be fit to one year of pre-participation and one year of post-participation billing data. However, these models are likely to give unstable or misleading results if changes take place at several points within these analysis periods. For commercial customers, especially larger customers, major changes can occur fairly frequently.
More complex models will also be explored, to account for all the changes identified during the year, and link to the metering data. The billing and metering data will also be used as a basis for calibrating the engineering models to actual consumption levels. Revised engineering estimates of savings will be based on these recalibrated models.

Task 7. Assess Factors Related to Savings Achieved and Free-Rider Effects: Because this evaluation will be a set of case studies, no explicit econometric model will be developed in this task, as is done for most of the other programs. Instead, the relationship of savings achieved to customer characteristics, including economic, attitudinal, and organizational, factors, will be explored qualitatively. On the basis of this assessment, a set of "without program" inputs to the recalibrated engineering models will be developed. These will be used to give a rough assessment of free-rider effects.

Task 8. Conduct Demand Savings Analysis: The demand analysis will be conducted according to the procedures outlined in Section 2. For the Nonresidential Customized Rebate program, the load data required will be obtained from the special load metering conducted under Task 1.

10.1 Program Description

The Industrial Process Optimization Program offers large commercial and industrial customers rebates for the installation of process-related energy efficiency improvements. Two quite distinct components comprise this program:

1. A direct rebate option for energy efficient, polyphase, 1800 RPM, NEMA Design B motors rated between 1 and 200 horsepower. Other motors (1200 RPM, 3600 RPM, or other sizes) and adjustable speed drives are also covered by this component, with pre-approval required.

2. A customized rebate option for industry- or site-specific process enhancements.

Eligible customers may either request a process energy assessment or apply directly for efficiency rebates. Energy assessments will be conducted by consultants.

The program includes both electric and gas energy efficiency measures in the combination service areas of Fort Dodge, Iowa City and the Iowa Quad Cities. Only gas energy efficiency measures are included in the gas-only service areas of Cedar Rapids and Ottumwa. Any package of process-related measures that can demonstrate energy and capacity savings may be eligible for a cash incentive.

Rebates are paid in two ways: prescriptive motor rebates are listed for 1800 RPM motors in the 1 to 200 horsepower range, and custom rebates are calculated for other motors, adjustable speed drives, and for process-related measures.. The motor incentives vary with motor size and are designed to reflect 50 percent of the participant's incremental capital cost.

The customized rebate amounts for electric and gas projects are calculated as the smallest of three amounts: the amount required to buy the payback down to two years, 80 percent of the avoided cost, or 50 percent of the incremental capital cost. Projects with less than a two year payback receive no rebate. Planning calculations showed that incentives would translate into about 38 percent of the participant's incremental cost for customized electric process measures and 50 percent for gas process measures.

3. The level of customers' interest in energy efficient motors and adjustable speed drives;

4. The nature of motor purchases, including:

   - *occasions for purchases*—burnout replacement versus adding new motors;
   - *delivery time requirements*—same day, next day, three to four days, or longer;
   - *usual purchasing practices*—order from previous supplier, inquire of more than one supplier, or put out to bid;
   - *importance of various decision factors*—first cost, delivery time, energy efficiency or operating costs, reputation of supplier, reputation of manufacturer, and reliability or quality; and
   - *existence of motor replacement plans;*
   - *criteria employed in deciding to rewind or purchase a replacement motor*; and

5. Ways to work most effectively with motor distributors and other suppliers.

Due to the varying needs of individual large customers, it is difficult to assess the market for specific process-related measures. Thus, the market assessment for process-related custom rebates should explore the frequency, timing and extent of process revisions with key customers and groups of customers. By assessing the magnitude of the naturally-occurring investment in revising and updating process equipment, Iowa-Illinois will develop a better sense of the appropriate level of effort for this program.

10.2.2 Process Evaluation

The process evaluation of the motors subprogram and the process optimization subprogram will address somewhat different questions. The evaluation of the motors subprogram will need to address a moderate number of utility and customer interactions with trade allies selling and servicing motors and drive components. The evaluation of the process optimization subprogram will need to address the details of a small number of transactions involving specific customers and vendors.

The evaluations of both subprograms will address the fundamental process evaluation questions outlined in Section 2.1.2. Because motor programs have met with varying success across the utility industry, additional effort will be focused on assessing customers' and trade allies' needs. Because the process optimization program relies heavily on individual interactions, questions regarding the timeliness and perceived quality of service will be included.

10.2.3 Impact Evaluation

Due to the small numbers of participants in both the motors and process-related custom measures subprograms, little general information about energy savings and demand reductions can be developed during the period covered by the first energy efficiency plan. Detailed information on the energy savings and demand reductions related to each specific installation will be developed by conducting individual case studies of each participant.

10.3 Technical Approach

10.3.1 Introduction

As noted above in the Evaluation Questions discussion, the two subprograms will require two distinct approaches. The motors subprogram will be addressed with methods appropriate for the goal of from seven to 34 participants per year. These methods will include semi-structured telephone or in-person interviews and the development of case studies. By the second and third year, when 20 or more customers are expected to participate, a sample of participants will be employed. The process subprogram components have goals of two to nine electric participants per year and one to five gas participants per year. Case studies with in-person interviews will be developed for each participant who will agree to a case study.

10.3.2 Market Assessment

The market assessment of the motors subprogram will involve semi-structured interviews with all participants and utility staff and small samples of two to three trade allies in each of Iowa-Illinois' electric operating areas. These interviews will provide qualitative answers to the questions listed above. Key questions will explore the nature of the market for energy-efficient motors and adjustable speed drives. An optional survey of a larger sample of customers could provide quantitative answers to the questions about the motor market and customers' motor purchases.

The market assessment of the process-related custom rebates will use data collected during regular visits with large customers. Iowa-Illinois keeps in close contact with major customers. These regular customer contacts will be used to inquire about customers':

- energy efficiency needs;
- goals for growth;
- plans for remodeling, renovation on facility additions, and
- plans for updating or replacing process equipment.

A brief, standardized questionnaire will be developed and employed whenever possible.

10.3.3 Process Evaluation

Information about customers' satisfaction with the process of participating in the Iowa-Illinois program will be obtained from the follow-up survey of all participants. Additional information about customers' satisfaction with energy efficient motors and process improvements will be obtained from successive interviews with participants who agree to be the subject of a case study.

10.3.4 Impact Evaluation

The evaluation of the Industrial Process Optimization program will be done as a set of case studies. The tasks are essentially the same as those for the Customized Rebate program, and are similar to those for the statistically based evaluations. In the Industrial Process Optimization program, motor drives, process heat, and other process energy uses will be addressed. These energy uses will not generally be considerations for the other programs. The appropriate energy-efficiency measures will be very customer-specific. As a result, the models required to estimated program effects, and the data required to support that modeling will also be very customer-specific.

The impact evaluation strategy for the tailored programs emphasizes on-site inspections. These inspections will provide accurate inputs for engineering calculations of savings. In addition, interval meters will be installed on motors or other major equipment predicted to be associated with large energy savings.

On-site visits will be conducted for all participants, or for a substantial fraction, depending on the number of participants. If all participants cannot be visited, priority will be given to customers with higher anticipated savings. Likewise, the end-use metering will target the customers and the specific pieces of equipment associated with the highest expected savings. Measures may also be targeted because there is greater uncertainty about their impacts, and resolving that uncertainty is important to decisions about the future program.

Task 1: Select the Cases for Study. Up to 10 participants per program year will be selected for case studies. If more than 10 customers participate in a given year, a non-random sample will be selected, based on the distribution of characteristics among the participants. Factors such as SIC, annual consumption and peak demand, and utility district will be considered in selecting the sample for study.

For each participant, the dates when energy efficiency measures are to be adopted will be determined. Some customers may participate in the program more than once, or may participate in multiple programs, for different measures. The multiple participation dates will be tracked.

Task 2: Collect Metering Data. Wherever possible, meters will be installed prior to the adoption of energy efficiency measures. End-use internal meters will be installed for motors and major process equipment and will operate both before and after new equipment is installed. Simple run-time meters will also be installed to collect hours of operation for motors and major process equipment.

Task 3: Conduct On-site Audits and Interviews. On-site audits and interviews will be conducted to determine detailed structural and operating characteristics of each case studied. Included in this data collection will be the key inputs required by the engineering models, as well as customer information required for the process evaluation.

Task 4: Refine the Engineering Models. This task will be performed as described in Section 2. Because each customer's circumstances, and the corresponding measures undertaken, will be very different, adjusting the model inputs and structure to reflect the situation accurately is essential. The impact evaluation for this program depends heavily on the accuracy of the engineering estimates.

Task 5: Estimate Savings Using Engineering Models and Conduct Implementation Analysis. This task will be performed as described in Section 2.

Task 6: Analyze Billing and Metering Data. For the Industrial Process Optimization, the end uses for which efficiency measures are adopted will ordinarily not be significantly temperature-dependent loads. As a result, degree-day models such as those described in Section 2 are unlikely to be of much use.

Instead, more complex models will be explored to account for all the changes identified during the year, and link to the metering data. The billing and metering data will also be used as a basis for calibrating the engineering models to actual consumption levels. Revised engineering estimates of savings will be based on these recalibrated models.

Task 7: Assess Factors Related to Savings Achieved and Free Rider Effects. Because this evaluation will be a set of case studies, no explicit econometric model will be developed in this task, as is done for most of the other programs. Instead, the relationship of savings achieved to customer characteristics, including economic, attitudinal, and organizational factors, will be explored qualitatively. On the basis of this assessment, a set of "without program" inputs to the recalibrated engineering models will be developed. These will be used to give a rough assessment of free-rider effects.

Task 8:. Conduct Demand Savings Analysis. The demand analysis will be conducted according to the procedures outlined in Section 2. For the Industrial Process Optimization program, the load data required will be obtained from the special load metering conducted under Task 1.

11.1 Program Description

The Iowa-Illinois Non-Residential New Building Construction Program provides financial incentives to encourage the incorporation of energy efficient practices in new commercial construction. Under this program two general types of incentives are offered:

- Design incentives to help defray the potentially higher cost associated with energy efficient design.
- Equipment incentives to reduce the first-cost barrier to include energy efficient measures in new buildings.

The program's equipment incentives are offered in two forms. The prescriptive portion of the program offers incentives on a per unit basis for a list of pre-screened new construction measures. The comprehensive program portion offers a $/kW incentive for savings calculated relative to a utility-defined baseline. Savings will be determined by computer modeling. For planning purposes, the incentive level for the custom program component was estimated to be, on average, $54/kW and is identical to that paid by the Non-Residential Building System Customized Rebate program. Actual rebate amounts will be calculated on a building and measure-specific basis. Measures with a payback of less than two years will not be eligible for an incentive. The prescriptive rebate amounts are based on an average incentive payment of $149/kW and are consistent with the Non-Residential Building Systems Shopping List Rebate Program. Actual incentives will vary by measure. These rebate levels represent an estimated 50 percent of the incremental measure cost under the prescriptive program pathway and 17 percent of the incremental cost for the comprehensive program component.

The program's participation and imapct goals for the first six years are shown in Table 11-1.

Figure 11-1
Non-Residential New Construction
Year 1-6 Goals and Estimated Impacts*

Prescriptive		1992	1993	1994	1995	1996	1997
	Participation Goal	0	4	9	14	17	19
	Energy Savings	0 MWh	122 MWh	396 MWh	823 MWh	1,342 MWh	1,921 MWh
	Demand Savings	0 kW Summer on-Peak	37 kW Summer on-Peak	121 kW Summer on-Peak	251 kW Summer on-Peak	410 kW Summer on-Peak	587 kW Summer on-Peak
Comprehensive	Participation Goal	0	0	2	2	4	4
	Energy Savings	0 MWh	0MWh	464MWh	929MWh	1,857 MWh	2,786MWh
	Demand Savings	0 kW Summer on-Peak	0 kW Summer on-Peak	140 kW Summer on-Peak	280 kW Summer on-Peak	560 kW Summer on-Peak	840 kW Summer on-Peak

* *Note: Participation figures show new participants for each year, Energy and Demand savings figures show the savings in each year for the cumulative number of participants.*

On a per building basis, the anticipated savings impact from the comprehensive component participants is several times larger than that of the prescriptive component participants.

11.2 Evaluation Questions

The difficulties entailed in evaluating a commercial new construction program are exacerbated by a number of factors, including, but not limited to:

- Lack of a preparticipation (before) consumption baseline.
- No building-specific preparticipation inventory data.
- For most comprehensive and some prescriptive-component participants, multiple measures installed per site.
- Small participant and nonparticipant populations and the resulting difficulties in generating a matched sample.
- Multiple decision makers regarding program entry.
- Long elapsed time from program application to measure installation verification.

While all these factors are not unique to a commercial new construction program, they will have a major effect in formulating the evaluation approach to the program and were considered in developing the questions listed below. Additional program evaluation issues which are common to all of the Iowa-Illinois programs are discussed in Section 2 and will be addressed as part of the evaluation methodology for the Nonresidential New Building Construction Program.

11.2.1 Market Assessment

- What is the current level of new construction in Iowa-Illinois' service territory? How do current economic conditions affect participation in the program?

- What are the characteristics of the participant and non-participant populations; building type and size, owner-occupied/tenant, spec-built/custom, urban/rural distribution?

- For the above characteristics, how do the comprehensive and prescriptive participants differ (Note that data on comprehensive program participants will be limited in this evaluation cycle)?

- What are the assumed baseline efficiencies for this program? How were they established and how well do they reflect current practice as well as participant practice in the absence of the program?

- Who is (are) the primary decision maker(s) regarding the installation of high efficiency measures and program participation?

- During the design/build process, when are the best points for the utility to influence the decision-making process?

- What is the typical design/build timeframe? How does the duration differ significantly by building type or size?

- How does the new construction program market differ from the replacement component of Iowa-Illinois' retrofit programs?

- Are there other equipment types that should be incorporated into the program.? If so, under the prescriptive or comprehensive path?

11.2.2 Process Evaluation

- What are the program's key goals? How were they addressed during program development?

- What important issues emerged during initial program start-up and implementation? Have program features or marketing approaches changed to address them?

11.2.3 Impact Evaluation

The impact evaluation of the Nonresidential New Construction Program will place primary emphasis on the use of engineering algorithms to compute energy and demand savings The use of billing analysis and metering will be limited, in large part because there is no pre- and post-treatment cases for each participant. Similarly, participant versus nonparticipant comparisons are of limited use given the small population involved. There is likely to be too much variation both within and between the two groups to allow valid comparisons to be made. Further, the use of billing analysis in new buildings may be constrained by numerous changes in occupancy over a short period, particularly in larger facilities built on speculation.

While the impact evaluation will rely primarily on engineering estimates, metering and other data collection activities will serve to verify and refine the inputs to the algorithms. Depending on the end uses to which the greatest program savings are ascribed, metering can be performed to verify and determine total runtimes and hours of use relative to Iowa-Illinois' avoided cost time periods. Additionally, metering can be useful to refine estimates of peak impacts at both the building and utility level.

Another important consideration in evaluating the program is whether there are significant lost opportunities among participants that are not captured by the program. Obtaining these DSM resources at a later time as retrofit measures will usually entail procuring them at a higher cost. Further, does the prescriptive or comprehensive program component obtain more of the available cost-effective DSM resources?

11.3 Technical Approach

11.3.1 Market Assessment

The majority of the market assessment tasks for the Nonresidential New Construction Program will be unique to the program, though some of the activities will be performed in conjunction with the other nonresidential programs, primarily the Customized Rebate Program and the equipment replacement components of the Shopping List and Process Optimization Programs. Two of the main goals of the both the realities of program operation as well as perceptions those involved with the program. The program's effectiveness be determined and recommendations made to improve prog performance.

The expected small participant and nonparticipant population during the program's first several years may restrict data colle efforts to the program's common elements. As a result, the of measure-specific issues addressed may be limited to those uses with significant activity, such as lighting.

The initial process evaluation task will be to review the progr history, development and current implementation. Both utili corporate and field staff will be interviewed. These interviews allow the evaluation team to provide immediate, short-term recommendations to the program's planning and implementa staff.

A related task will be a review of the program's tracking data and application forms to insure that they are configured to colle appropriate program data. As participants enter the program 1993 and 1994, the tracking system will again be reviewed to of the required program data are properly entered.

Participant and nonparticipant surveys will be done in conjunc with the market assessment task. The entire participant popula will be interviewed as well as a similar, or possibly larger, sampl nonparticipants. However, as noted above, the participant sampl projected to be quite small. Only four participants are anticipated 1993 and eleven participants in 1994 (nine prescriptive and two comprehensive). One of the important new construction progra issues is: *Who makes the decision to participate?* As a result, multip surveys may be done for each facility. Targeted parties would be owner, developer and the design professionals involved in equipment specification.

Prior to the participant and nonparticipant surveys, focus groups semi-structured personal interviews will be held with owners, developers, design professionals and trade allies. The goal of the focus groups will be to provide more immediate program feedback and to provide direction on the development of the participant/ nonparticipant surveys or interview guides.

11.3.3 Impact Evaluation

The measurement of energy and demand savings among program participants will rely primarily on engineering algorithms for which selected inputs have been validated by both telephone surveys and on-site data collection. The on-site data collection activities will consist of facility surveys and targeted end-use metering. The initial tasks are similar, if not identical, to those for the Nonresidential Building System Shopping List Rebates (NR-1) Program and are repeated below. Metering data from the shopping list program will be used to supplement that collected from the new building construction program participants.

Task 1: Define Wattage Assumptions for Installed and Replaced Lighting Technologies. Wattage assumptions will be developed for the installed technologies offered under NR-1 and for the baseline technologies identified in the implementation plans. To ensure consistency, XENERGY recommends that wattage assumptions be obtained from a common data source such as the American National Standards Institute (ANSI).

Task 2: Estimate the Full Load Hours (FLH), Demand Factor (DF), Coincidence Factor (CF), and Interaction Factors (IF). FLH, DF, IF, and CF will be obtained from secondary data for the first year evaluation. For subsequent years, end-use metering data will be collected for a sample of participants to provide seasonal full load hours and coincidence factors for each of the building sectors. Run meters will be installed to verify the demand factor assumptions and the participant survey will be used to determine the percentage of space conditioned floor space (see Validation Plan). These data will be supplemented by available utility load research data.

Task 3: Develop Engineering Algorithms. The engineering algorithms for lighting, air conditioning, heat pumps, set-back thermostats and other installed measures will be developed to produce preliminary saving estimates.

Task 4: Conduct Billing Analysis. The billing histories for program participants and a sample of nonparticipants will be obtained from Iowa-Illinois' customer billing files. The data will be both screened to remove any anomalies and weather normalized to

On-Site Surveys and Metering. Through the third quarter of 1994, all program participants up to a total of 20 will receive on-site visits. The purpose of these visits will be to:

- Further characterize the participant population.
- Verify measure installation and determine if the measure is performing as intended.
- Assess the extent of measure capture and identify lost opportunities in other end uses.
- Examine operating schedules to confirm impact algorithms assumptions.
- Install run-time and other metering equipment to collect data on end-use operating patterns and demand impacts.

A.1 Guiding Principles

Metering represents one component of an overall evaluation strategy that incorporates several tools, including surveys, statistical analysis, and engineering methods. Many tools are necessary for accurate evaluations because energy or demand savings can never be measured or observed directly. Rather, savings must be derived from differences in measured and observed quantities. The consequences of this constraint severely limit the effectiveness of using metered data alone to formulate DSM savings estimates. However, metered data analysis, when combined with other evaluation techniques, can certainly increase the precision of DSM savings estimates.

The extent to which metering constitutes a cost-effective evaluation tool for a given DSM program will be driven by several factors, including:

- The number and variety of program participants
- The types of measures involved
- The nature and magnitude of the savings per measure
- The availability and reliability of supplemental data acquired using other evaluation tools

A partial list of tools available for program or technology evaluation includes the following:

Monitored Information

- visual inspections
- customer surveys
- specification and quality control of tracking data
- obtaining manufacturer's data (from nameplate data and documents)
- engineering or simulation models

Metered Information

- whole building demand or energy metering
- end-use kVA, kW, and/or kWh metering
- end-use operating hour metering
- measurement of end-use variables such as cooling load, temperature, motor speed, equipment efficiency, etc.

We can define monitored information as that which can be obtained without physical measuring equipment (e.g., the monitored items shown above.) Similarly, we can define metered information as that which requires physical measurement equipment of varying sophistication and cost (i.e. the metered items shown above). Whether or not additional monitoring or metering is justified in a given situation depends on whether or not the value of the increased precision obtainable from a given method exceeds the method's cost and whether another method exists which could provide the same information for less cost.

All of the tools listed above could help to improve the precision of energy and demand savings estimates, but the cost required to obtain the different forms of information varies greatly. The ultimate goal is to combine all of these tools in a cost-effective way to ensure that a satisfactory level of precision is attained. The targeted level of precision should be dictated by its associated costs and benefits (discussed in Section A.4).

Two new developments improve the role of metering and monitoring within DSM evaluation. The first is hardware-based. The metering industry has developed high volume/low cost metering techniques which record the components of energy savings, such as hours of operation (component of energy savings are discussed in Section A.3).

The second development, which is software-based, consists of new analytical techniques being developed which permit the computation of precision using *component*-based metering and monitoring. This analytical technique allows an evaluator to identify what data components in an impact evaluation have the most influence on the precision of savings estimates. These components can then be targeted for additional data collection (discussed in Section A.3).

This section describes the steps that will be followed to formulate a metering and monitoring plan for the Iowa programs. The following principles will be used to guide the development of this plan:

- Evaluation goals will be well defined before data collection begins.
- The incremental cost of data collection should not exceed the incremental value of metering.
- The plan will use multiple evaluation tools designed to capture the variables of interest (see list above for a partial list of these tools).

A key objective of each plan will be to define cost-effective metering. Traditionally, gross savings estimates have been computed using one or a combination of methods, including engineering estimates, statistical billing analysis, and end-use or facility metering. The cost and resulting precision obtained from these methods vary significantly. Typically, the relative cost/precision of each of these methods for a single site observation is as follows:

	Cost	Precision
Engineering Estimates	Low	Low
Statistical Analysis	Med/Low	Med
Metering	High	High

Intuitively, any move up the unit cost scale from engineering estimates to metering must be justified by the value of the resulting precision gains. Ideally, the most cost-effective precision gains can be attained by proficiently combining all three methods for an entire sample. Identifying the components of the value of precision will be discussed in Section A.4.

A.2 Sources of DSM Impact Uncertainty

The precision of a program savings estimate is driven by the following underlying factors:

- data collection limitations
- baseline energy-use assumptions
- variance in participant characteristics and behavior

Data collection limitations introduce error in the form of instrumentation error, observation error, normalization error and temporal error. Each of these error types is discussed below:

> Instrumentation error refers to the precision limitation of physical measurement equipment. This error is usually random and negligible compared to other error sources (e.g., on the order of 1% or less for most measurement devices).
>
> Observation error refers to mistaken interpretation and/or inaccurate reporting of observed data, such as inaccurate equipment counts or mis-read model numbers. The significance of observation error varies, depending on the data collection mechanism used. For example, error tends to be large when customers are asked to estimate equipment counts over the telephone. Conversely, the error tends to be small when trained auditors conduct on-site surveys.
>
> Normalization error is introduced when data are adjusted using regression techniques to account for weather, occupancy, and behavioral changes over time and across participant and nonparticipant populations. This component can contribute significantly to savings estimate error.
>
> Temporal error refers to inaccuracies in projected long-term savings estimates based on extrapolated data gathered over a limited period of time. This error can also constitute significant error in impact evaluations.

Savings estimates for a single metered participant contain error due to both data collection limitations and baseline energy use assumptions (the latter being defined as the changes in usage which would have occurred in the absence of an energy-savings measure). To further complicate the problem, error terms, such as instrumentation error, become magnified when applied to savings estimates. This results from the fact that savings cannot be directly measured, but must be computed as the *difference* between two measured quantities. Since the instrument error is given as a percentage of the quantity being measured, rather than a percentage of this difference, the error quantity constitutes a larger percentage of the savings estimate.

Program savings estimates for all participants contain, in addition to data collection and baseline error, error due to variance in participant characteristics and behavior. When estimating program savings, inferences about the population of participants are drawn from sample data. Although attempts are made to adjust the data to reflect variations in physical and behavioral customer characteristics, significant error can result. The size of this error depends on many things, including the magnitude of customer to customer variance, the number of observations drawn from the population, and biases in sample selection.

By itself, metering can provide precise information about individual customers. However, metering alone cannot give precise information for the whole customer population, unless the population is very homogeneous or the metering is conducted for a very large number of customers.

Before specifying a metering and monitoring plan, all of the components of evaluation error should be assessed to ensure that the robustness of the plan is compatible with the overall evaluation approach. The targeted precision of the data collection plan, including metering, should be combined with the targeted precision of the other error components to ensure that the targeted precision of the overall evaluation can be attained. Generally, the metering and monitoring plan should be designed so as to not exceed the "error budget" allocated to this evaluation component. Conversely, the plan design should not aim for extremely high precision, if these errors will be negligible compared to the targeted precision of the other evaluation components. This consideration is of special concern if the cost required to attain such precision negatively affects the DSM program's cost-effectiveness.

A.3 Estimating Metering and Monitoring Error

As discussed earlier, energy and demand savings cannot be measured directly, but must be calculated from measured values. Therefore, for a given metering and monitoring plan, some type of savings equation is required to establish a savings estimate. Although these equations may resemble engineering algorithms, evaluation techniques using actual metered or observed data should not be confused with "engineering algorithm" approaches, which typically use generic measure savings data combined with limited program-specific data such as number of participants.

One approach to identifying the information that should be gathered as part of a metering and monitoring plan is to investigate the error contributions of the individual components that make up these savings calculations. Estimating the achievable precision improvements for these different components can help indicate which are the most likely candidates for additional monitoring or metering. This concept can best be illustrated with an example:

A simplified estimate of savings resulting from a lighting efficiency improvement for a given participant can be calculated using the following equation:

$$E = (W_b - W_a) \times HRS \times INST \qquad \text{(Equation 1)}$$

Where: E = Est. energy savings due to lamp replacement project
W_b = Est. wattage of inefficient lamp
Wa = Est. wattage of efficient lamp
HRS = Est. operating hours (before & after usage assumed equal for simplicity)
INST = Est. number of installations

As shown in the equation above, the energy savings estimate is derived from several variables. Consequently, the savings estimate precision is a function of the individual precision of each of these variables. One could estimate the expected value and standard deviation of each of these variables based on information from manufacturer's data, program tracking databases, metered samples, and other sources. In some instances where explicit information is unavailable, it may be more practical to estimate the absolute confidence interval for a given variable rather than a standard deviation. For the purposes of our example, assume that the following estimates of the expected value and expected error were obtained from various sources (for error estimates, assume 90% confidence levels):

Table 1

Variable	Expected Value	Expected Error
W_b	177 W	5%
W_a	112 W	5%
HRS	3000 hrs	50%
INST	300 lamps	20%

For a given confidence level, the approximate degree to which the uncertainty of any one of these individual variables affects the precision of the total savings estimate can be calculated using established methods. The results for a 90% confidence level are shown below:

Table 2

Variable	Expected value of E	Contribution to Precision	% Contribution to Total Variance
W_b	58.5 MWh	± 13.62%	5.87%
W_a	58.5 MWh	± 8.62%	2.35%
HRS	58.5 MWh	± 50.00%	79.12%
INST	58.5 MWh	± 20.00%	12.66%
Total	58.5 MWh	+ 56.21%	100.0%

The overall precision of the savings estimate can then be calculated by applying a root-sum square-formula to the individual uncertainty values attributable to each variable. The result in our example provides an overall savings precision of ±56.21% with 90% confidence.

If we wanted to improve this estimate of precision, it is apparent from the table above that reducing the uncertainty of operating hours (variable HRS), if possible, would provide us the most significant benefit since this variable contributes most significantly to the uncertainty of our overall savings estimate. If it were possible to reduce the uncertainty of this variable by half (i.e., 25%), the resulting precision of the savings estimate would be improved to ± 35.8%. Conversely, in the context of our example, less benefit would be gained by reducing the uncertainty of the variable Wb. Even if it were possible to reduce this uncertainty to 0%, the total savings precision would only improve to about ± 54.5%. Note that an underlying assumption of our example is that the rated wattage of the pre-measure fixtures was available in our tracking database. If not, the error of this variable would be significantly greater and would probably constitute most of the savings estimate error. This demonstrates the value of gathering monitored data, or visually observed data, as well as metered data. With careful quality control, reasonably accurate monitored data can usually be gathered for much less cost than metered data.

As illustrated in this example, the goal of a data collection plan should be to improve the estimates of all the savings equation variables in a cost-effective manner to attain the error budget established for the data collection component of the evaluation. In our example, a large number of run-time hour observations could be obtained for the same cost as would be required to collect a few observations using 15-minute interval demand metering equipment. As a result, greater precision could be obtained for less cost.

The analysis framework illustrated in the example above can be generally applied to any DSM program to identify what data components should be gathered to most efficiently improve precision. An important component of such an analysis is estimating and comparing the cost of attaining measured observations of the various data components.

A.4 Metering Technology

Large-scale metering of DSM measures for impact purposes first began as a utility load research activity. In this activity utility companies investigated how their customers used energy in order to better design rate structures and to better predict load growth.

Under PURPA regulations, utility companies are required to meter a representative sample of customers using whole-building fifteen minute data collection. Typically, the samples are selected by rate class and are used to support rate filings. Commercial building PURPA samples are in the 300 to 500 range.

Since 1980, several utility companies have undertaken ambitious enduse metering programs. The goal of these programs has been to account for all the electric use in buildings by end use at fifteen minute intervals. In the Bonneville Power ELCAP program, about 65 commercial buildings have been metered by end use since the mid-1980s. Southern California Edison and Central Maine Power have metered hundreds of residential facilities (hot tubs are an SCE end use). Pacific Gas and Electric's commercial enduse metering program is about two years old and includes 45 samples.

All of the end-use metering programs mentioned above share certain characteristics including high expense, intensive data collection, large amounts of data. As one practitioner has described it, data arrive with the force and volume of water through a fire hose. However, the process has provided fundamental insight into the electrical consumption patterns of buildings.

Because of the history of metering in load research, the first approach to metering for evaluation purposes has been in that mold, with an orientation towards comprehensive, long-term, end-use data collection. However, because of the fundamental difference in purpose between load research and evaluation, the mold can be broken.

Fundamentally, load research requires an understanding of all electrical consumption in a building; DSM evaluation requires an understanding of a technology. DSM evaluation must capture the performance of a technology and the performance variance from building to building.

A.4.1 The New Menu of Metering Options

Metering does not come in one flavor anymore. The metering industry has recently developed a number of devices which are specifically designed to capture components of the savings equation at a considerably lower cost then classic end-use metering. Figure 1 presents the range of metering technologies available. A description of the metering characteristics is provided below:

Target. Load research projects have targeted end uses. These projects attempt to account for all the energy consumed in the building by end use. For DSM evaluation, the target is usually an appliance. An appliance can be a motor, a fixture, a lighting circuit, a chiller or other device which is examined for its energy performance.

Period of Metering. The duration of the metering period can be characterized as follows:

- Long-term - Approximately three months or more.
- Short-term - Approximately less than three months.
- Spot - Instantaneous readings.

The split between long and short term is somewhat variable. Long-term metering is characterized by the need for a modem for transmission of data and a requirement for purchase, rather than lease of metering equipment due to the extended installation period.

Frequency. The frequency at which the data are logged is characterized as:

- Hourly - Data recorded at least hourly for load profiles. May include daily recording for some applications.
- Period Logging - Accumulated run hours per time period (i.e., off-peak and on-peak) over the period of metering.
- Totalized - Accumulated run hours or kWh over the period of metering.

Measured Value. Typically, measurements include:

- Volts - Voltage is generally constant. Measurement requires an electrician using devices which make conductor contact.
- Amps - Electric current varies with load. Measurement of electric current alone provides a good indicator of part-load performance of a device. Measurement can be non intrusive using a current transducer which looks like a big clothes pin. The measurement usually requires an electrician.
- kW - Complete power analysis can require seven sensors: voltage and current for each leg of a three phase circuit and a separate single sensor for power factor.
- On-Off - For static load devices (lights, motors), hours of operation is an important parameter. Hours of operation can be measured by wiring a device into the circuit of interest and monitoring the on-time. This requires an electrician and is intrusive. A second approach is to indirectly sense on/off time. Lighting hours of operation can be sensed using an optical pick-up which counts on-hours when light is sensed. Motor hours of operation can be tallied using a device which counts on-hours when a magnetic field is sensed. These devices are mounted with magnetic strips and do not require electricians.
- Other inputs - Instrumentation is available to measure temperature, pressure, flow, rmp and virtually any required parameter - at a cost.

Resulting Parameter. The resulting data include:

• Hours of Operation	Accumulated equipment run-time.
• kWh	Accumulated electric consumption.
• Operating Profile	Percent of full load at hourly intervals.
• Part Load	Percent of full load at which the device is operating.
• Power Analysis	Complete load characterization including power factor, amps, volts.
• kW profile	kW consumption at hourly intervals

Data Processing. Administration, computers, QC and data massaging can add substantially to the cost of metering beyond installation and hardware. Long-term hourly metering requires a layer of administration to QC, manage and massage the data. Short-term hourly metering can be installed, deinstalled and analyzed by a single engineer; however, analysis generally requires a PC and number crunching. Devices with LED readouts (amp meters, hours of operation loggers) require a simple read which can be taken by any staff and requires little additional analysis.

The final column presents a cost per point. The cost is somewhat incremental. It is interesting to note that the lower end costs can be driven primarily by the cost of getting on site. If some of these measurements can be integrated with verification surveys, low costs are indicated.

A.5 Suggested Development Approach Toward Metering and Monitoring Plans

Suggested steps for development of a detailed metering plan are:

1. Identify cost-effectiveness criteria
2. Quantify measure parameters
3. Use gross impact methodology to derive sample sizes
4. Examine measures and develop unit costs for metering
5. Iterate to select most cost-effective mix

A.5.1 Identify Cost-Effectiveness Criteria

A framework should be developed to evaluate the cost-effectiveness of metering and monitoring. This framework requires the utility to assess the value of the information gained in DSM evaluations. Improving savings estimate precision alone does not *directly* or *immediately* increase the actual kW and kWh savings resulting from the DSM measure itself. Improving precision simply provides the utility with improved accounting controls. These accounting controls constitute DSM "overhead" in that they do not contribute to the savings "revenue" generated by DSM programs. To realize any benefit from these accounting controls, a utility must actively apply them to improve its operations. Still, benefit is associated with these accounting controls and it is these benefits we need to identify and *quantify* in order to complete an optimization model for planning impact evaluations.

Several functional areas within a utility could potentially benefit from having detailed information on measures, but uses of this information could vary substantially. Some potential benefits of improving the precision of DSM impacts include:

1. Increased reliability of the DSM "resource" as part of a utility's integrated resource plan, thereby reducing the reliance on costly and environmentally harmful supply-side options.
2. Improved planning and implementation of DSM programs as a result of improved inputs to technical assessment and DSM market research.
3. Improved reporting to regulatory bodies, resulting in increased DSM incentive earnings and in favorable and expeditious treatment.

4. Risk avoidance attributed to the small, diverse nature of DSM resources as compared to the large, centralized nature of traditional supply-side options.

To assign value to these various benefits, valuation models can be developed based on a given utility's unique circumstances. To develop such models, a utility should survey the areas within its organization that could potentially benefit from this information and quantitatively interpret how each of these areas measures or perceives the value of this information. This exercise can reveal surprising applications of impact evaluation data and possibly locate additional sources of funding to support evaluation work.

To demonstrate this concept, one justification for increasing the precision of evaluation estimates might be to verify DSM's contribution to a utility's generating capacity (base-load or peaking). To maintain a reliable margin of reserve capacity, resource planners may only be willing to incorporate the most conservative estimate of DSM savings into the utility's integrated resource plan (IRP). Therefore, improving the reliability, hence precision, of DSM savings could increase DSM's avoided-capacity value. This relationship, graphically depicted in Figure 2, could form the basis of a model for establishing the avoided-capacity value of precision increases in DSM evaluation:

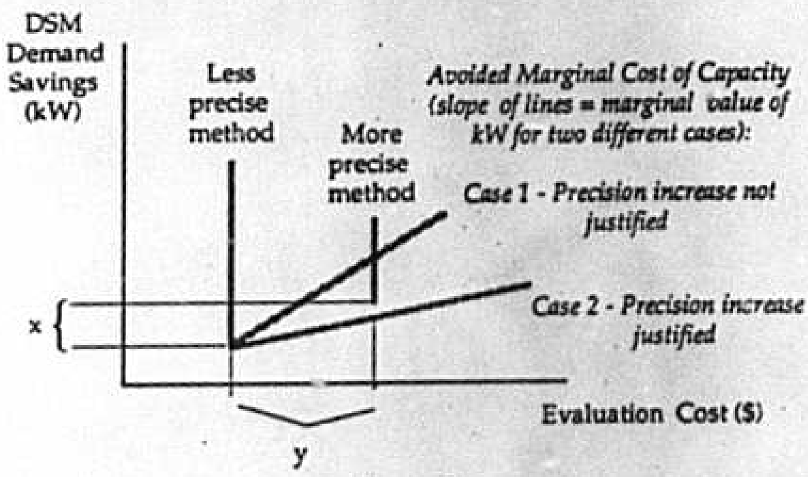


Figure 2

Estimated avoided-capacity savings for two different evaluation approaches vs. cost

In the figure, the vertical solid lines represent the error bounds on DSM savings estimates for two different evaluation approaches. The cost of the more precise method exceeds that of the less precise method by amount *y*. However, the more precise method yields a savings estimate with a narrower error bound than that of the less precise method. Therefore, if the more precise measure were used, resource planners might be able to include more DSM capacity in their integrated resource plan, thereby allowing them to defer generating capacity equal to amount x. The gray lines in Figure 2 represent two very different avoided-cost scenarios (i.e., the utility's marginal cost of procuring additional generating capacity). In the first scenario (Case 1 in Figure 2), for the incremental cost of the more precise method, the amount of generating capacity that the utility could have procured is greater than amount x. Therefore, the more precise evaluation method would not be cost-effective. In the second scenario (Case 2 in Figure 2), for the incremental cost of the more precise method, the amount of generating capacity which the utility could procure is less than amount x. Therefore, the more precise evaluation method might be justified. However, because all we have compared is the relative cost-effectiveness of the two methods, another method and/or sample size might exist which provides even greater cost-effectiveness.

A.5.2 Quantify DSM Measure Parameters

Information from research on measures must be prioritized, i.e., a value placed on the information in order to rationally allocate precious metering resources. Metering provides better information about a measure performance. Better information about measures that are expected to be significant contributors to DSM programs is intuitively more valuable then poor information about an insignificant measure.

Therefore, the next step is to examine programs and identify for each of the measures in the program:

- Projected impact;
- Projected utility benefit (in avoided cost, incentive, as decided above);
- Number of participants;
- Cost-effectiveness;
- Cost to install: customer, incentive and administration costs;
- Key impact assumptions (hours of operation, setting, etc.); and
- Other issues, such as political sensitivity, problems, etc.

A.5.3 Examine Gross Impact Method to Determine Sample Sizes

The strategic gross impact methodology will determine the number of sites which must be monitored. Since technology characteristics are measured, rather than program characteristics, the plan will identify sample sizes by measure rather than by program.

Strategies of metering and monitoring can vary, depending on their specific role with the larger evaluation framework. Common strategies are described below:

- *Case studies/Technology Assessment* - Metering can be used as a tool of fundamental building science research. The purpose of a case study is to characterize the performance of a technology, not to compute program-wide impacts. The results of a case study are not statistically significant; however they may lead to conclusions about cost-effectiveness or provide estimates for planning purposes. They can be useful in understanding implementation problems, in determining the technology performance (though not the variability of that performance from site to site) and as marketing fodder.

The number of metered sites is largely based on judgment. For a technology demonstration site, one site may be adequate. For gross scaling of estimates, a site in each major segment may be adequate.

Metering for a case study, rather than for direct impact evaluation, may require additional metering points to further characterize performance of a system. For example, a kWh totalizer operating over the pre and post periods may be adequate for impact calculations of a VSD pump application. For technology assessment, additional points such as pump speed would be useful for development of load curves.

Some examples: XENERGY conducted a metering case study of a daylight dimming system for Northeast Utilities. The results from a single site demonstrated that technology could achieve significant savings and a difference in performance between manufacturers. Metered points included incident insolation, fifteen minute kW of test and control, fifteen minute ambient temperature.

A case study might be useful for analyzing a new or controversial technology such as liquid pressure amplifiers, a heavily marketed device which is claimed to reduce refrigeration energy use.

In a case study for NYSERDA, Xenergy will demonstrate the efficacy of occupancy sensor controls in multiple space types in four buildings using fifteen minute kW metering.

- *Direct Impact Measurement* - Impacts can be directly measured by metering a sample or a census of participants. Any sample selection must be statistically significant in order for this approach to provide valid results for the program as a whole. The sample size will be driven by the site-to-site variability of the parameter of interest.

  Usually, it is not cost-effective to meter a statistically significant portion of program participants, unless there are very few participants.

  Direct impact measurement is most cost-effective for large impact measures in large C&I programs. It is reasonable in these cases to establish a protocol requiring before and after measurement.

  Some Examples: Boston Edison Company requires baseline and post-installation metering of participants in their large C&I Energy Partnership Program. Too few sites have gone to implementation as of now to gage the success.

Green Mountain Power has applications for industrial measures which are predicted to account for 1-2% of total impact (all programs). Each of these measures will be metered.

Bay State Gas will meter a statistically significant sample of participants in their DSM programs, using whole building fifteen minute data. However, the metering is accomplished through the use of an automated billing system which greatly reduces the incremental cost of metering.

- *Model Calibration* - Metering can be used to calibrate or validate engineering or statistical models. As an example, engineering estimates of hours of operation for lighting can be improved by metering a sample of participants. The ratio of estimated savings to actual savings can be applied to all participant estimates resulting in a better estimate. Typically, the samples required for model calibration are smaller than the samples required for direct impact measurement.

A.5.4 Develop Unit Costs for Metering

After identifying the sample sizes, the next consideration is the unit metering cost. This cost is determined by the measures and parameters of interest. Sample sizes and targeted measures and parameters are defined in the previous step. Using the unit metering costs from Section IV, initial data collection costs can be estimated. Unit metering costs can be by program implementation considerations. For example, some of the low-cost metering techniques, such as light loggers or amp-readings, can be implemented inexpensively if built into the DSM program delivery. The cost of traveling to and from the site can cost more than the technology itself. For example, if account representatives will be on-site as part of the DSM program, the representative could mount light loggers during the visit. Other parameters may be more easily obtained from the customer, such as full load amp readings. In many cases, the installation contractor can be requested to provide this type of information as part of the DSM program agreement.

A.5.5 Iterate

After initial estimating costs and sample sizes, cost-effectiveness must re-assessed. The mix of sample sizes and metering technologies can be readjusted to improve cost-effectiveness.

B.1 Engineering Algorithms

The Iowa-Illinois tracking system will be used to provide preliminary savings estimates. A series of engineering algorithms will be specified and estimated to measure program impacts on an on-going basis. The routine tracking of the program's performance will indicate areas of the program that should be modified to enhance its savings impacts. Iowa-Illinois Non-residential Building Systems Shopping List Rebate Program (NR-1) provides monetary incentives to commercial and industrial (C&I) customers who install energy efficient lighting, air conditioning, heat pumps, and programmable thermostats.

The following engineering algorithms represent the preliminary model specifications to calculate winter kW, summer kW, and peak and off-peak kWh impacts. Equations 1 to 3 apply to interior lighting technologies. Equations 4 to 6 apply to air conditioners and the algorithms for heat pumps are shown in equations 7 to 9.

B.2 Interior Lighting Algorithms

(1) $\text{Summer kW Impact} = SAVE \times NUM \times CF_S \times DF \times (1 + FAC \times WCB) \times (1 - FRF) \times (1 - SBF) \times PLF$

(2) $\text{Winter kW Impact} = SAVE \times NUM \times CF_W \times DF \times (1 - FHT \times WHP) \times (1 - FRF) \times (1 - SBF) \times PLF$

(3) $\text{Energy Impact} = SAVE \times NUM \times DF \times [(1 - FHT \times WHP_E) \times (FLH_{WOP} + FLH_{WP}) + (1 + FAC \times WCB_E) \times (FLH_{SP} + FLH_{SS} + FLH_{SOP})] \times (1 - FRF) \times (1 - SBF) \times ALF$

B.3 Space Conditioning Rebate Program Algorithms

The algorithms apply to air conditioning x technologies.

(4) $\text{Summer kW Impact} = (1 / EER_{BASE} - 1 / EER_{HIEFF}) \times 12 \times SIZE \times CF_S \times DF \times (1 - FRF) \times (1 - SBF) \times PLF$

(5) $\text{Winter kW Impact} = (1 / EER_{BASE} - 1 / EER_{HIEFF}) \times 12 \times SIZE \times CF_S \times DF \times (1 - FRF) \times (1 - SBF) \times PLF$

WHP	=	Weighted heating penalty. Similar to the cooling bonus, heating penalty accounts for the fact that more efficient lighting generates less heat. This additional heat must be supplied by the heating system. The penalty is based on the efficiency of the heating system.
WCB_E	=	Weighted energy cooling bonus. The cooling bonus only occurs when cooling and lighting are on at the same time. The summer demand impact algorithm assumes that at system peak, both are on. However, during the entire year this is not always the case so an energy cooling bonus is used.
WHP_E	=	Weighted energy heating penalty. Similar to the cooling bonus, heating penalty only occurs when lighting and heating are on at the same time.
SIZE	=	Unit Size in Tons.
CF_S	=	Summer coincidence factor. It is calculated as the ratio of the end-use demand at the hour of system summer peak to the peak demand for the end use.
CF_W	=	Winter coincidence factor. It is calculated as the ratio of the end-use demand at the hour of system winter peak to the peak demand for the end use.
DF	=	Demand factor. It is the ratio of peak demand of an end use to the installed capacity for the end use.
$SEER_{BASE}$	=	Base Cooling Seasonal Energy Efficiency Ratio (SEER).
$SEER_{HIEFF}$	=	New Cooling SEER.
$HSPF_{BASE}$	=	Base Heating Seasonal Performance Factor (HSPF).
$HSPF_{HIEFF}$	=	New HSPF.
EER_{BASE}	=	Base Energy Efficiency Ratio (EER).
EER_{HIEFF}	=	New EER.
12	=	Factor converting tons to MBtuh.

FRF	=	Free ridership factor. The percentage of participants who would have installed the energy-efficient equipment without the program.
SBF	=	Snapback factor. The percentage of participants who increase their usage after installing the energy-efficient equipment.
PLF	=	Transmission and distribution losses at peak. It is calculated as peak generation divided by peak demand at the meter.
ALF	=	Average T&D loss factor. It is calculated as total annual energy generation divided by total energy demand at the meter.
FLH_{SP}	=	Summer peak full load hours.
FLH_{SS}	=	Summer shoulder full load hours.
FLH_{SOP}	=	Summer off peak full load hours.
FLH_{WP}	=	Winter peak full load hours.
FLH_{WOP}	=	Winter off peak full load hours